Kati Kristiansson

Genetics of Cardiovascular Disease:

A Candidate Gene Study of *USF1*

Publications of the National Public Health Institute A 8/2008

Department of Molecular Medicine
National Public Health Institute
and
Department of Medical Genetics
University of Helsinki

Helsinki, Finland 2008

Kati Kristiansson

GENETICS OF CARDIOVASCULAR DISEASE:

A CANDIDATE GENE STUDY OF *USF1*

A C A D E M I C D I S S E R T A T I O N

To be presented with the permission of the Medical Faculty,
University of Helsinki, for public examination in the Small Lecture Hall,
Haartman Institute, on April 11th, at 12 noon.

National Public Health Institute, Helsinki, Finland

and

Department of Medical Genetics, University of Helsinki, Finland

Helsinki 2008

Publications of the National Public Health Institute
KTL A8 / 2008

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 978-951-740-787-8
ISSN 0359-3584
ISBN 978-951-740-788-5 (pdf)
ISSN 1458-6290 (pdf)

Yliopistopaino
Helsinki 2008

Supervised by

Academy Professor Leena Peltonen-Palotie
National Public Health Institute,
Department of Molecular Medicine and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland

Docent Markus Perola
National Public Health Institute,
Department of Molecular Medicine and
University of Helsinki,
Department of Medical Genetics
Helsinki, Finland

Reviewed by

Docent Katariina Kainulainen
Department of Medicine, Division of Infectious Diseases
Helsinki University Central Hospital
Hospital District of Helsinki and Uusimaa
Helsinki, Finland

Docent Katarina Pelin
Department of Biological and Environmental Sciences
Division of Genetics
University of Helsinki
Helsinki, Finland

Opponent

Professor Juni Palmgren
Mathematical Statistics
Medical Epidemiology and Biostatistics
Stockholm University
Karolinska Institutet
Stockholm, Sweden

Kati Kristiansson, Genetics of Cardiovascular Disease: A Candidate Gene Study of *USF1*
Publications of the National Public Health Institute, A8/2008, 111 Pages
ISBN 951-740-787-8; 951-740-788-5 (pdf-version)
ISSN 0359-3584; 1458-6290 (pdf-version)
http://www.ktl.fi/portal/4043

ABSTRACT

Cardiovascular diseases (CVD) are major contributors to morbidity and mortality worldwide. Several interacting environmental, biochemical, and genetic risk factors can increase disease susceptibility. While some of the genes involved in the etiology of CVD are known, many are yet to be discovered. During the last few decades, scientists have searched for these genes with genome-wide linkage and association methods, and with more targeted candidate gene studies.

This thesis investigates variation within the upstream transcription factor 1 (*USF1*) gene locus in relation to CVD risk factors, atherosclerosis, and incidence and prevalence of CVD. This candidate gene was first identified in Finnish families ascertained for familial combined hyperlipidemia, a common dyslipidemia predisposing to coronary heart disease. The gene is a ubiquitously expressed transcription factor regulating expression of several genes from lipid and glucose metabolism, inflammation, and endothelial function.

First, we examined association between *USF1* variants and several CVD risk factors, such as lipid phenotypes, body composition measures, and metabolic syndrome, in two prospective population cohorts. Our data suggested that *USF1* contributes to these CVD risk factors at the population level. Notably, the associations with quantitative measurements were mostly detected among study subjects with CVD or metabolic syndrome, suggesting complex interactions between *USF1* effects and the pathophysiological state of an individual.

Second, we investigated how variation at the *USF1* locus contributes to atherosclerotic lesions of the coronary arteries and abdominal aorta. For this, we used two study samples of middle-aged men with detailed measurements of atherosclerosis obtained in autopsy. *USF1* variation significantly associated with areas of several types of lesions, especially with calcification of the arteries.

Next, we tested what effect the *USF1* risk variants have on sudden cardiac death and incidence of CVD. The atherosclerosis-associated risk variant increased the risk of sudden cardiac death of the same study subjects. Furthermore, *USF1* alleles

associated with incidence of CVD in the Finnish population follow-up cohorts. These associations were especially prominent among women, suggesting a sex-specific effect, which has also been detected in subsequent studies.

Finally, as some of the low-yield DNA samples of the Finnish follow-up study cohort needed to be whole-genome amplified (WGA) prior to genotyping, we evaluated whether the produced WGA genotypes were of good quality. Although the samples giving genotype discrepancies could not be detected before genotyping with standard laboratory quality control methods, our results suggested that enhanced quality control at the time of the genotyping could identify such samples. In addition, combining two WGA reactions into one pooled DNA sample for genotyping markedly reduced the number of discrepancies and samples showing them.

In conclusion, *USF1* seems to have a role in the etiology of CVD. Additional studies are warranted to identify functional variants and to study interactions between *USF1* and other genetic or environmental factors. This *USF1* study, and other studies with low DNA yield of some samples, can benefit from whole genome amplification of the low-yield samples prior to genotyping. Careful quality control procedures are, however, needed in WGA genotyping.

Keywords: USF1, cardiovascular disease, atherosclerosis, genetic association analysis, follow-up study, lipid metabolism, familial combined hyperlipidemia, metabolic syndrome, cardiovascular risk factors, whole genome amplification

Kati Kristiansson, Genetics of Cardiovascular Disease: A Candidate Gene Study of *USF1*
Kansanterveyslaitoksen julkaisuja, A8/2008, 111 sivua
ISBN 951-740-787-8; 951-740-788-5 (pdf-versio)
ISSN 0359-3584; 1458-6290 (pdf-versio)
http://www.ktl.fi/portal/4043

TIIVISTELMÄ

Sydän- ja verisuonitaudeilla (SVT) on vakavia sosiaalisia ja terveydenhuollollisia vaikutuksia eri yhteiskunnissa. Yleisiä sairastumisen riskiä lisääviä tekijöitä tunnetaan useita ja ne voivat olla ympäristöllisiä, perinnöllisiä tai osittain molempia. Vaikka useita verenkiertoelinten sairauksiin vaikuttavia perinnöllisiä tekijöitä on jo tunnistettu, monia sairastumiseen vaikuttavia geenialueita on vielä tunnistamatta. Näitä alueita tutkijat ovat viime vuosikymmenten aikana etsineet koko perimän laajuisilla kytkentä- ja assosiaatioanalyyseillä, sekä tarkemmin kohdennetuilla ehdokasgeenianalyyseillä.

Tässä väitöskirjatyössä analysoitiin upstream transcription factor 1 (*USF1*) -geeniä, tutkien sen yhteyttä SVT riskitekijöihin, ateroskleroosiin, sekä SVT:n ilmaantumiseen ja esiintymiseen. Tämä geeni tunnistettiin alun perin suomalaisissa perheissä, jotka valittiin familiaalisen kombinoidun hyperlipidemian (FKH) tutkimusta varten. USF1 säätelee rasva- ja sokeriaineenvaihduntaan, tulehdus-reaktioihin ja endoteelin toimintaan liittyvien geenien ilmentymistä, mikä tekee siitä mielenkiintoisen ehdokasgeenin sydän- ja verisuonitaudeille.

Tutkimme ensin miten perinnöllinen vaihtelu *USF1*:n alueella on yhteydessä perinteisiin SVT:n riskitekijöihin, kuten rasva-arvoihin, lihavuusmuuttujiin, sekä metaboliseen syndroomaan. Vaihtelu *USF1*:n alueella liittyi näihin riskitekijöihin kahdessa väestöpohjaisessa seurantaotoksessa. Lähes kaikki yhteydet havaittiin tarkastelemalla seuranta-otosten erityisryhmiä, kuten sydänsairaita, tai metabolista syndroomaa sairastavia. Tämä viittaisi monimutkaisiin yhdysvaikutuksiin *USF1*:n toiminnassa.

Seuraavaksi tarkastelimme *USF1*:n vaikutusta ateroskleroosin muodostumiseen suonitasolla. Käytössämme oli kaksi äkkikuolleista miehistä koostuvaa suomalaista ruumiinavausaineistoa. Aineisto sisälsi tietoa erityyppisistä ateroskleroosimuuttujista ja havaitsimme *USF1*:n yhdistyvän varsinkin verisuonten kalkkeutumiseen.

Tutkimme vielä *USF1*:n vaikutusta sydänperäisen äkkikuoleman riskiin sekä sydän- ja verisuonitaudin ilmaantumiseen. Havaitsimme ateroskleroosiin liittyvän riskialleelin kasvattavan myös sydänperäisen äkkikuoleman riskiä samassa tutkimusaineistossa. Lisäksi *USF1*:n eri muodot vaikuttivat SVT:n riskiin suomalaisissa seurantaotoksissa. Seurantaotoksissa riskivaikutus ilmeni pääasiassa

naisilla, mikä viittasi sukupuolille ominaiseen geenivaikutukseen, jota on jälkeenpäin ehdotettu myös muissa tutkimuksissa.

Testasimme myös miten koko perimän laajuinen DNA:n monistus (WGA) vaikuttaa genotyypityksen luotettavuuteen. Tarkoituksena oli tarkastella niitä suomalaisen väestöotoksemme näytteitä, joiden DNA saanto oli niin vähäistä, että niiden monistaminen ennen genotyypitystä oli tarkoituksenmukaista. Osa WGA-näytteistä antoi genotyypityksessä ristiriitaisia tuloksia, eikä näitä näytteitä voitu havaita muiden näytteiden joukosta normaaleissa näytteiden laaduntarkkailuissa ennen genotyypitystä. Genotyypitystulosten tarkasteluissa useat ristiriitaiset näytteet erottuivat kuitenkin muista hyvin. Lisäksi kahden erillisen WGA-reaktion yhdistäminen yhdeksi genotyypitettäväksi näytteeksi paransi genotyypityksen luotettavuutta huomattavasti.

Tässä väitöskirjatyössä saimme siis viitteitä *USF1*-geenin roolista verenkiertoelinten sairauksissa. Tulevat *USF1*-tutkimukset sisältänevät selvitystyötä geenialueen toiminnallisista varianteista, sekä tutkivat tarkemmin geenin yhdysvaikutuksia sukupuolen, sekä muiden ympäristöllisten ja perinnöllisten tekijöiden kanssa. Tässä *USF1*-tutkimuksessa, kuten myös muissa vähäsaantoista DNA:ta sisältävissä tutkimuksissa, voidaan hyödyntää näytteiden koko perimän DNA:n monistusta ennen genotyypitystä. WGA-näytteiden genotyypitys vaatii kuitenkin huolellisen laaduntarkkailun, jotta genotyypitysvirheitä tuottavat näytteet tunnistetaan ennen tilastollisia analyysejä.

Avansanat: USF1, sydän- ja verisuonitaudit, ateroskleroosi, geneettinen assosiaatiotutkimus, seurantatutkimus, rasva-aineenvaihdunta, familiaalinen kombinoitu hyperlipidemia, metabolinen syndrooma, sydäntaudin riskitekijät, koko perimän DNA:n monistus

CONTENTS

ABBREVIATIONS

APOB	apolipoprotein B
APOE	apolipoprotein E
ATPIII	Adult Treatment Panel III
BMI	body mass index
bp	base pair
CAC	coronary artery calcium
CHD	coronary heart disease
CI	confidence interval
CRP	C-reactive protein
CVD	cardiovascular disease
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
F11R	F11-receptor
FCHL	familial combined hyperlipidemia
GWA	genome-wide association
HDL	high density lipoprotein
HR	hazard ratio
HSDS	Helsinki Sudden Death study
ICD	international classification of diseases
IL6	interleukin 6
kb	kilobase
LD	linkage disequilibrium
LDL	low density lipoprotein
LDLR	low density lipoprotein receptor
Mb	megabase

MDA	multiple displacement amplification
MetS	metabolic syndrome
MI	myocardial infarction
mRNA	messenger-RNA
OR	odds ratio
PAD	peripheral arterial disease
PCR	polymerase chain reaction
RNA	ribonucleic acid
SCD	sudden cardiac death
SD	standard deviation
SMC	smooth muscle cells
SNP	single nucleotide polymorphism
T2DM	type 2 diabetes mellitus
TC	total cholesterol
TG	triglycerides
ULSAM	Uppsala Longitudinal Study of Adult Men
USF1	upstream transcription factor 1
WGA	whole-genome amplification

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles referred to in the text by their Roman numerals:

I Auro K, **Kristiansson K**, Zethelius B, Berne C, Lannfelt L, Taskinen MR, Jauhiainen M, Perola M, Peltonen L, Syvänen AC. *USF1* gene variants contribute to metabolic traits in males in a longitudinal 32-year follow-up study. *Diabetologia*. 2007 Dec 21.

II **Kristiansson K**, Ilveskoski E, Lehtimäki T, Peltonen L, Perola M, Karhunen PJ. Association analysis of Allelic Variants of *USF1* in Coronary Atherosclerosis. *Arterioscler Thromb Vasc Biol.* In press.

III **Komulainen K**, Alanne M, Auro K, Kilpikari R, Pajukanta P, Saarela J, Ellonen P, Salminen K, Kulathinal S, Kuulasmaa K, Silander K, Salomaa V, Perola M, Peltonen L. Risk alleles of *USF1* gene predict cardiovascular disease of women in two prospective studies. *PLoS Genet.* 2006 May;2(5):e69.

IV Silander K, **Komulainen K**, Ellonen P, Jussila M, Alanne M, Levander M, Tainola P, Kuulasmaa K, Salomaa V, Perola M, Peltonen L, Saarela J. Evaluating whole genome amplification via multiply-primed rolling circle amplification for SNP genotyping of samples with low DNA yield. *Twin Res Hum Genet*. 2005 Aug;8(4):368-75.

Publication I also appears in the thesis of Kirsi Auro (2007).

1 INTRODUCTION

Cardiovascular diseases (CVD), which involve the heart, brain and peripheral circulation, affect millions of people worldwide, and are among the leading causes of death and disease burden both in high-income and low-income countries (Lopez et al. 2006). Along with several environmental and biochemical risk factors, disease susceptibility is modified by a genetic predisposition (Stegmayr et al. 1997; Kuulasmaa et al. 2000; Flossmann et al. 2004; Yusuf et al. 2004; Yarnell et al. 2005; Goldstein et al. 2006; Mayer et al. 2007). This genetic background is evidenced by familial aggregation of the disease, and family and twin studies have proposed that genetic factors explain 17 to 61% of the observed variation in CVD mortality or morbidity in the study populations (Bak et al. 2002; Zdravkovic et al. 2002; Fox et al. 2003; Fischer et al. 2005; Moskau et al. 2005; Wienke et al. 2005).

Understanding how genetic variation in individuals and populations contributes to the biological pathways and mechanisms leading to CVD has been a subject of an intense research. Prospects for successful gene identification have been hindered by the complex nature of the disease; whereas in monogenic disorders variation(s) in a single gene locus with high penetrance results in a disease and the Mendelian segregation of the risk allele is observable in pedigrees, common complex diseases such as CVD are characterized by low penetrance, locus heterogeneity, epistasis, and variation in the disease phenotype (Lander and Schork 1994; Risch 2000).

Appropriate selection of genetic markers and study samples can greatly aid dissection of the genetic background of a complex disease. The utility of isolated populations, such as Finland, is not necessarily limited to the mapping of rare diseases, but complex disease research can also greatly benefit from their features (Wright et al. 1999; Heutink and Oostra 2002; Service et al. 2006). Furthermore, increased information on single nucleotide polymorphisms (SNPs) and their haplotype-tagging properties in public databases ("The International HapMap Project" 2003), as well as improved multiplexing in SNP genotyping platforms have facilitated research (Robinson and Speed 2007; Steemers and Gunderson 2007). Methods for amplifying DNA at the whole genome level have also appeared (Hawkins et al. 2002), and we evaluated the utility of one of these amplification methods in SNP genotyping of a population based CVD study sample.

Several genome-wide linkage or association studies have tried to identify genomic regions harboring genes contributing to CVD (Pajukanta et al. 2000; Francke et al. 2001; Broeckel et al. 2002; Harrap et al. 2002; Chiodini and Lewis 2003; Hauser et al. 2004; Helgadottir et al. 2004; Wang et al. 2004; Samani et al. 2005; Wang et al.

2005; Engert et al. 2007; Larson et al. 2007; Matarin et al. 2007; Nilsson-Ardnor et al. 2007; O'Donnell et al. 2007; Samani et al. 2007). Additionally, numerous small-scale and large-scale candidate gene studies have examined the effect of various genes from the lipid and inflammation related pathways on CVD morbidity and mortality. In this work, we studied a particularly interesting candidate gene for CVD; the upstream transcription factor 1 (*USF1*), originally implicated in the etiology of familial combined hyperlipidemia (FCHL) (Pajukanta et al. 2004). This gene encodes a ubiquitously expressed transcription factor, which participates in the regulation of the expression of several genes from lipid and glucose metabolism pathways (Pajukanta et al. 2004; Naukkarinen et al. 2005). The chromosomal location of the gene also coincides with a linkage peak from several type 2 diabetes mellitus (T2DM) studies (Hanson et al. 1998; Elbein et al. 1999; Vionnet et al. 2000; Hsueh et al. 2003; Ng et al. 2004a; Xiang et al. 2004) and from a metabolic syndrome study (Ng et al. 2004b). These disorders, along with FCHL, are strongly associated with CVD (Goldstein et al. 1973b; Tuomilehto et al. 1996; Austin et al. 2000; Isomaa et al. 2001; Lakka et al. 2002; Hopkins et al. 2003; Malik et al. 2004; Ninomiya et al. 2004; Huxley et al. 2006).

2 REVIEW OF THE LITERATURE

2.1 The role of genetics in the etiology of disease

Two major breakthroughs in the field of genetics in the mid-20th century were the determining of the three-dimensional structure of the DNA molecule by Francis Crick, James D. Watson, Rosalind Franklin, and Maurice Wilkins in 1953 (Watson and Crick 1953) and the solving of how triplet mRNA codons specify the twenty amino acids through the combined efforts of scientists in the 1960s (Crick et al. 1961; Nirenberg and Matthaei 1961; Matthaei et al. 1962; Khorana 1968). Later, the first findings of disease-causing genes began to emerge: The discoveries of human oncogenes in 1981 were among the first reported (Perucho et al. 1981). The easy amplification of DNA by the polymerase chain reaction (PCR) (Mullis et al. 1986) in the 1980s greatly facilitated genetic research and the sequencing projects of eukaryotic genomes finally led to a successful completion of the Human Genome Project in 2003 (Lander et al. 2001; Waterston et al. 2002; International Human Genome Sequencing Consortium 2004). Together, these advances have enabled the present-day efforts of dissecting the genetic background of various diseases.

The role of genetics greatly varies depending on the trait or disease. While in some diseases, such as in many diseases of the Finnish disease heritage (Norio 2003), genes are the sole determinants of illness, in others, such as some virus or bacterial infections, the impact of genes is trivial or absent and the illness is determined by the interaction of genes and environmental factors, or by environment alone. Furthermore, the gene defect(s) causing the disease might be acquired during life and be present in only few cells, like in some cancers, or they might be inherited and be present in all cells of the body. In addition, a disease can arise due to a chromosomal abnormality, a single mutation in a gene, or several mutations in many genes.

Whatever the mechanisms leading to the genetic diseases are, elucidating their nature and biology is of great importance. Understanding how a genetic disorder arises, offers tools for genetic counseling, gene-therapy, pharmacogenomics, and medicine, which attempt to prevent and treat these diseases.

2.2 Strategies for identifying genes underlying complex traits

Many genes responsible for monogenic (Mendelian) diseases have been identified in the past two decades by linkage analysis and positional cloning methods. Most diseases and traits, however, involve multiple genes and environmental influences and their genetic dissection has proved to be more challenging (Table 1).

Table 1. Complicating factors in genetic studies of complex diseases

Genetic factors
Unknown mode of inheritance
Unknown allele frequencies
Genetic heterogeneity
Epistasis
Gene-environment interactions
Phenotypes
Difficulties in diagnosis and classification of the phenotype
Variable expressivity of the phenotype
Incomplete penetrance
Late onset of diseases
Pleiotropy
Phenocopies
Technical issues
Completeness and usability of genetic data in public databases
Affordable large-scale genotyping methods
Statistical issues
Limited statistical power
Multiple testing
Other
Publication bias

In monogenic disorders, variation(s) in a single gene locus with complete or high penetrance results in a disease, and the Mendelian segregation of the risk allele is observable in pedigrees. Distinguishing the genetic background contributing to a complex disease is more difficult (Lander and Schork 1994; Risch 2000); the onset of the disease is the result of several genes functioning together with environmental risk factors. The overall contribution of each gene can be small, the frequencies of the risk alleles vary from rare to common, and epistasis, locus heterogeneity, pleiotropy, and low penetrance further complicate the gene discovery. Variation in the disease phenotype also obscures its genetic research; complex diseases can vary

in many instances, such as the age of onset, severity of symptoms, and in their etiological mechanisms.

To minimize the effect of any factors complicating gene discovery, strategies for identifying genes underlying complex traits should include careful consideration of the most appropriate study samples, genetic markers, and analysis techniques depending on the disease or phenotype analyzed (Lander and Schork 1994; Risch 2000; Glazier et al. 2002).

2.2.1 Genome and its markers

The studies of the inheritance of human disease were facilitated with the recognition that DNA sequence of the genome contains naturally occurring variation (Petes and Botstein 1977; Botstein et al. 1980). The later sequencing of the human genome (Lander et al. 2001; Venter et al. 2001; International Human Genome Sequencing Consortium 2004) further accelerated disease gene discovery by providing invaluable information about the loci and nature of various kinds of polymorphic genetic markers (Table 2).

The human genome is estimated to contain around 20,000 protein coding genes, whose transcripts cover about 1.9% of the euchromatic genome (International Human Genome Sequencing Consortium 2004). In addition to protein coding exons, gene loci comprise introns, promoters, and other regulatory regions. Additionally, the end product of some genes is a RNA molecule, which is not translated into a protein. Instead, those RNA molecules act as transfer-RNA, ribosomal-RNA, or have regulatory functions. Mutations in these gene-containing DNA sequences arise spontaneously or are induced by mutagens and they can impair the normal function of the gene, which can occasionally lead to a less favorable or a beneficial phenotype. If these mutations occur in germline cells, they are passed on to offspring and create variation in the gene pool of the population, providing material for natural selection, genetic drift, and other evolutionary forces (Tishkoff and Verrelli 2003). Although changes that occur outside of the gene loci do not interfere with the gene function, they can be utilized in genetic research as markers for the more functionally relevant genetic variation (Palmer and Cardon 2005).

Table 2. Examples of variation in the human genome

Variation	Features	Size
Restriction fragment length polymorphisms (RFLP)	Restriction enzymes cleave DNA at specific sequences producing fragments of DNA. Variation of the DNA sequence between individuals at the restriction site can be detected as differing fragment lengths in electrophoresis.	Restriction sites from 4 to 12 bp
Single nucleotide polymorphisms (SNP)	A single nucleotide of the DNA sequence differs in comparison with a reference sequence.	1bp
Microsatellites	Tandem arrays of repeat units	Repeat units 1-6 bp length, copy number of repeats usually 10-30
Minisatellites	Tandem arrays of repeat units	Repeat units 8-100 bp length, copy number of repeats from 5 to over 1,000
Structural/Copy number variation	Insertions, deletions, translocations, inversions, and duplications of large segments of DNA sequence. Copy number variants are present at a variable copy number in comparison with a reference genome.	From 1 kb to more than 3 Mb
Aneuploidy	A change in the number of chromosomes	Whole chromosomes

Insertion and deletion (INDEL) polymorphisms of the genome overlap several of the categories presented in this table; INDEL polymorphisms range from 1 to 10,000 bp in length and include expansions of 1-15 bp repeat units (Mills et al. 2006).
Modified from Jobling et al. 2004 and Feuk et al. 2006.

Small-scale variation in the DNA sequence involves one or few nucleotides, whereas large-scale mutations, such as duplications, inversions, deletions, and translocations, can involve large segments of the genome (Table 2). During the last few decades, the types of variations most rigorously studied have changed as the techniques for detecting genomic variation have evolved.

Restriction fragment length polymorphisms were among the first markers used in disease gene discovery (Botstein et al. 1980), and they have also been widely used in paternity testing and in genetic fingerprinting in forensic science. Genetic research was further facilitated with the recognition of microsatellite markers (Weber and May 1989), which consist of repeating units of 1-4 base pairs in length and which have been especially useful in identifying disease-causing gene loci by linkage analysis in pedigrees. In present-day research, single nucleotide polymorphisms (SNP) and lately also copy number variation, have gained extensive attention (Landegren et al. 1998; Sebat et al. 2004; Redon et al. 2006).

The human genome harbors over 11 million SNPs (over 6 million of which are validated) (dbSNP Build 128, accessed 28.11.2007), which most often display only

two different alleles. The effect of a SNP on the gene function varies depending on the location and nature of the SNP. For instance, an exonic SNP may alter the amino-acid sequence of the gene product and lead to impaired protein function, and a SNP residing in a promoter region of the gene may alter the rate of transcription. In genetic research, the detection of the functional SNP is greatly aided by genomic linkage disequilibrium: A new mutation arises in a DNA sequence where specific SNP alleles exist. This combination of SNP alleles, a haplotype, is then inherited as a unit from the parental genomes, unless a recombination event in meiosis breaks the linkage between the SNP alleles. Between closely located SNPs, a recombination event is rare and genetic linkage keeps specific SNP alleles together; as a result, haplotype-tagging SNPs that are screened in genetic association studies provide information about several other SNPs in the loci also (Johnson et al. 2001; Gabriel et al. 2002; "The International HapMap Project" 2003).

Single nucleotide polymorphism studies have conventionally screened SNPs at one or several particular loci, perhaps after a whole-genome microsatellite screening has first identified this candidate region. In the last few years, however, whole-genome association scans using SNP arrays have also become feasible (Carlson et al. 2004; Hirschhorn and Daly 2005).

Despite intense research, discovery of complex disease or phenotype associated SNPs has proved to be challenging and it has been suggested that effective research should also utilize larger scale structural variation of the genome (Sebat et al. 2004; Redon et al. 2006; Stranger et al. 2007). These copy-number variations include gains or losses of large segments of genomic DNA, and they may function in the etiology of a disease for example through altered gene dosage or disruption of gene-coding or regulatory sequences. Well-known CVD related genes with copy number variation include thrombomodulin and lipoprotein a (Pollex and Hegele 2007).

When the objective is to identify new candidate loci for traits and diseases, a genome-wide linkage or association scan of genetic markers is the ideal study design. Often this is the case when the biological pathways leading to the disease are not known. If the molecular mechanisms involved in the etiology of the disease are established, however, then candidate genes can be examined for a marker-phenotype association (Tabor et al. 2002). Candidate genes and loci are also inspected after the initial genome-wide studies to identify functional variants of the region and to study the effect of the identified disease-causing variants at the population level.

2.2.2 Choosing the study sample

In order to minimize the effect of factors obscuring gene discovery, careful consideration should be put into the selection of the optimal study sample in research of common, complex diseases. An important issue is the sample size, which has to measure up to the estimates from power calculations for the study (Risch and Merikangas 1996). The study sample also has to match the aim of the study, whether it is to find new candidate genes for a disease or to study previously identified genes more thoroughly. Replication of the findings in other, independent study samples is crucial (Lander and Kruglyak 1995), although the reasons for lack of replication can sometimes be related to heterogeneity between studies rather than to spurious initial associations (Sillanpää and Auranen 2004).

Family-based study samples are often used in the earliest stages of a complex disease gene search (Burton et al. 2005). Familial aggregation of a disease, measured by recurrence risk ratios or correlation measures, provides indirect evidence of an existing genetic contribution, although non-genetic factors can also add to the aggregation (Lander and Schork 1994; Burton et al. 2005). Estimates of heritability, the proportion of phenotypic variance that is attributable to genetic variation among individuals, of a trait can be obtained for instance from twin studies (Lander and Schork 1994; Burton et al. 2005). Segregation analysis in families have traditionally examined whether genes with a large enough effect for their identification exist (Lander and Schork 1994; Burton et al. 2005). When evidence of a genetic influence has been obtained, the genomic location of specific genes contributing to the disease is screened for with linkage analysis (Dawn Teare and Barrett 2005). While linkage analysis may identify regions of the genome segregating with the disease in pedigrees, further fine-mapping studies of the regions are required to identify causative genes and variants. This is achieved with association analysis, linkage-disequilibrium mapping, or haplotype analysis, either in family-based study samples or in population-based study samples (Cordell and Clayton 2005).

Due to improvements in genotyping platforms, population-based samples without information on family relations can nowadays also be used in genome-wide SNP association scans for disease-associated genomic regions (Carlson et al. 2004; Hirschhorn and Daly 2005). In contrast to a genome-wide approach, if the molecular mechanisms behind the disease are known, genetic variation at candidate gene loci can be tested for disease association in case-control samples (Tabor et al. 2002). Population-based case-control studies circumvent some of the problems arising in family-based studies. For instance, diseases with late onset are more easily studied with a case-control approach where the availability of DNA usually is not a limiting

factor. Conversely, false positive associations due to multiple testing are a major problem in genome-wide SNP scans of case-control samples (Storey and Tibshirani 2003; Carlson et al. 2004). Bias introduced by population stratification is an additional difficulty complicating case-control studies (Marchini et al. 2004).

When disease-associated genetic variants have been identified with family-based or case-control studies, their impact on the morbidity of the general population can be examined in cohort studies (Manolio et al. 2006). In prospective cohort studies participants are usually examined at baseline, after which they are followed for a time. Often participants are also invited for examination at different time points during the follow-up period. The incidence of the disease during the follow-up is recorded and its association with genetic variation is studied. Cohort studies are not ideal for rare diseases since a large number of participants is needed to gain enough cases for analysis; for those diseases a case-control approach is more appropriate (Cardon and Bell 2001; Manolio et al. 2006).

The selection of controls in cohort studies is less prone to bias than in case-control studies, since those who have not taken ill during the follow-up represent controls in the analysis (Manolio et al. 2006). A case-cohort study is a special type of a cohort study where data is analyzed only from those who developed the disease of interest during the follow-up and from a randomly selected subcohort from the original follow-up sample (Figure 1). This type of study was originally designed as an alternative way of conducting large survey studies where it is too expensive and time consuming to analyze all data from all participants.

The effect of the disease-causing genetic variation at the cellular level can be examined for example with gene expression microarrays and proteomics (Lockhart and Winzeler 2000; Sellers and Yates 2003). The aim in these studies is often to investigate whether a specific polymorphism affects the amount or activity of the gene's mRNA, or whether an effect on the protein product of the gene occurs. Expression analyses can also reveal disease-causing genes through abnormal patterns of protein expression in those who have the disease (Lockhart and Winzeler 2000; Sellers and Yates 2003).

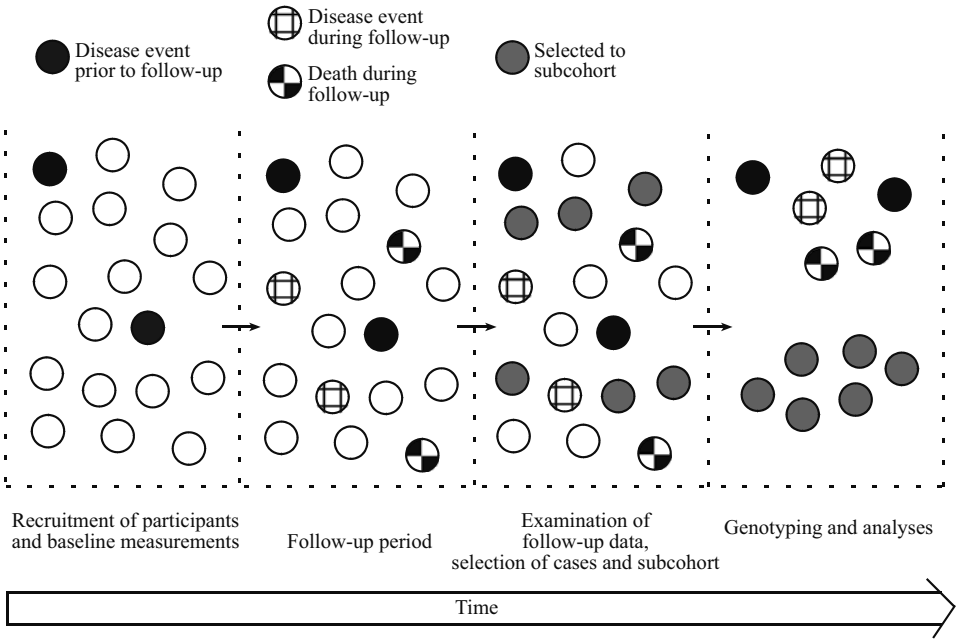

Figure 1. A case-cohort study. After a follow-up, cases of interest are identified from the original follow-up cohort and are selected for genotyping and analyses. A random subcohort is also selected from the follow-up cohort to be used as a comparison group in the time-to-event analysis. These subcohort members may include some cases, as cases are eligible for the subcohort selection as well. This way, the subcohort represents the original follow-up cohort.

In all study designs, a prudent choice of study population can reduce the genetic and environmental heterogeneity hampering the discovery of complex disease genes (Wright et al. 1999; Peltonen et al. 2000). Population isolates represent a more homogenous population to sample from, and they have more than proven their value in studies of Mendelian diseases (Peltonen et al. 1999). Features of these special populations, such as Finland, can facilitate genetic research of complex diseases as well: Uniform environmental and lifestyle-related risk factors may aid the detection of genetic risk factors contributing to the disease in reasonable-sized study samples (Peltonen et al. 2000). Furthermore, isolates are more likely to possess the same set of predisposing genes and alleles due to founder effect, genetic bottlenecks, and/or reduced migration (Wright et al. 1999; Peltonen et al. 2000). Population stratification, which may lead to false gene-disease associations in case-control studies, is also less likely to be present (Shifman and Darvasi 2001; Cardon and Palmer 2003). Another important advantage of population isolates is increased linkage disequilibrium (LD) present in their genomes (Service et al. 2006). Until resequencing of the entire gene loci and their surroundings becomes feasible for

large study samples, a gene-disease association can only be observed by genotyping a causative variant or a variant in high LD with it. Genome-wide association (GWA) scans in population isolates consequently require fewer genotyped markers, or capture more genomic variation through haplotype blocks with the same amount of genotyped markers, than GWA scans in outbred populations (Service et al. 2006). High LD can, on the other hand, complicate the identification of a causative gene variant and fine-mapping studies might therefore require the utilization of more outbred populations (Jorde 2000).

2.2.3 Amount of DNA as a limiting factor in studies

In addition to other genetic and environmental factors, genetic research is sometimes complicated by the amount of DNA available for the study. In most genetic studies the DNA to be genotyped is extracted from whole blood samples. While DNA yield from the blood varies with different storage conditions and DNA extraction methods, also natural variation between individuals occur (Cushwa and Medrano 1993; Nederhand et al. 2003; Alanne et al. 2004). If this variation is dependent on the health status of an individual as suggested (Alanne et al. 2004), selection of samples based on the amount of DNA available can introduce a bias to the study from the beginning.

Whole genome amplification of low-yield DNA samples prior to genotyping can provide a solution for preventing a DNA amount related bias (Hawkins et al. 2002). Several amplification methods exist, including degenerate oligonucleotide primed PCR (DOP–PCR) (Telenius et al. 1992; Cheung and Nelson 1996), primer extension preamplification (PEP) (Zhang et al. 1992; Dietmaier et al. 1999), and multiple displacement amplification (MDA) (Dean et al. 2001; Lovmar and Syvänen 2006). The MDA method (Figure 2) has stood out in comparison with other amplification methods providing good genomic coverage, high call rates, and reliable genotypes, while still having the lowest amplification bias (Lovmar and Syvänen 2006). Utilizing this method in genetic studies would aid the genotyping of otherwise non-genotypable low-yield DNA samples. Systematic quality control should be, however, conducted in order to identify possible spurious genotypes produced.

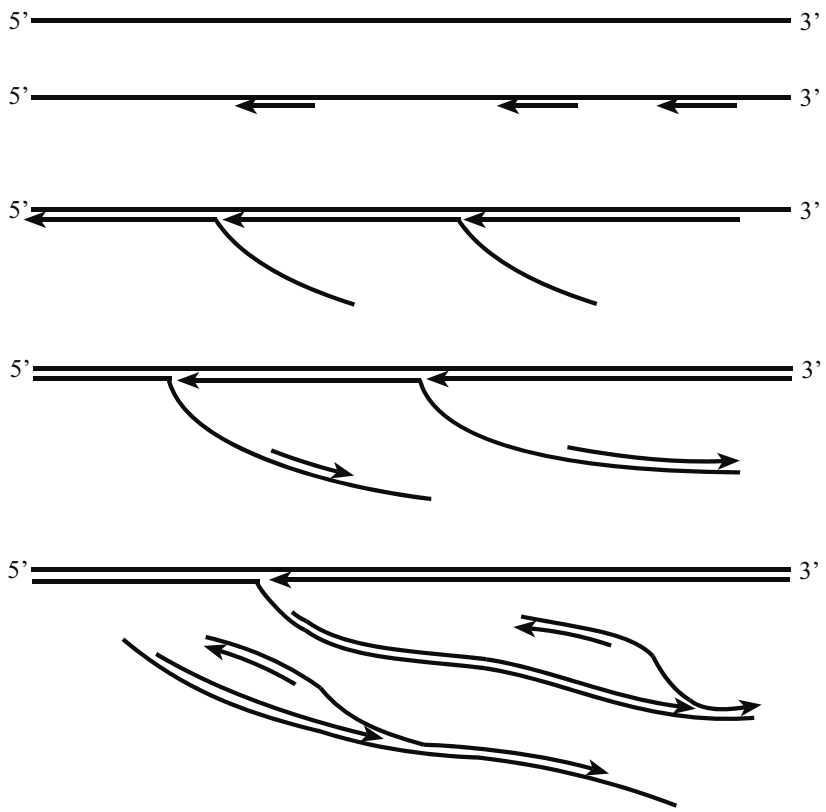

Figure 2. Whole genome amplification by the multiple displacement amplification (MDA) method. Random hexamers act as primers, annealing to complementary sequences in the single-stranded DNA molecule. The primers are elongated by DNA polymerase activity and in the process the 5’ ends of the elongated strands are displaced by upstream strands growing in the same direction. These displaced strands serve as templates for new priming events in the opposite direction. As the new priming events and strand displacements take place, an expanding hyperbranched network is formed, generating a massive amount of copies of the original DNA. Modified from Lage et al. 2003 and Lovmar and Syvänen 2006.

2.3 Overview of atherosclerosis and cardiovascular disease

Common, complex forms of cardiovascular disease, such as coronary heart disease and stroke (Table 3), present a great health burden on modern societies. In both high-income and low-income countries, cardiovascular diseases were among the leading causes of death and disease burden in 2001 (Lopez et al. 2006). In 2005, the

statistical database of the Finnish National Cardiovascular Disease Register (Mähonen et al. 2000) recorded 1743 coronary events (I20-I25) per 100,000 inhabitants in the age group of 25-74. The number of patients and deaths due to ischemic stroke (I63) was 309 per 100,000 inhabitants. First signs of these common diseases appear as early as in the childhood (Berenson et al. 1998; Strong et al. 1999).

Table 3. Cardiovascular diseases relevant to this thesis, according to the ICD-10 classification

Diseases of the circulatory system (I00-I99)
Ischaemic heart diseases I20-I25
Angina pectoris I20
Acute myocardial infarction I21
Subsequent myocardial infarction I22
Certain current complications following acute myocardial infarction I23
Other acute ischaemic heart diseases I24
Chronic ischaemic heart disease I25
Cerebrovascular diseases I60-I69
Cerebral infarction I63
Diseases of arteries, arterioles and capillaries I70-I79
Atherosclerosis I70
Aortic aneurysm and dissection I71
Arterial embolism and thrombosis I74

Cardiovascular diseases usually result from a complex series of events affecting arterial blood vessels; atherosclerosis. Atherosclerotic plaque formation begins with the accumulation of lipoprotein particles, such as low density lipoprotein (LDL) and lipoprotein a (Lp(a)), in the intima layer of the arteries (Figure 3). These lipoproteins transmigrate to the intima through the endothelium at sites of lesion predilection, which are determined by characteristics of the blood flow in the arteries. When the circulating levels of these lipoproteins are raised, the rate of their accumulation in the subendothelial matrix is also increased. In the intima, the lipoproteins undergo oxidation and other modifications, such as aggregation, lipolysis, and proteolysis. These modifications subsequently stimulate the endothelium to produce other molecules involved in the plaque formation. High-density lipoprotein (HDL) has an anti-atherogenic role in these initial processes: it removes excess cholesterol from the peripheral tissues and also inhibits lipoprotein oxidation. In response to the accumulation of modified lipoproteins in the vessel wall, endothelial cells produce

pro-inflammatory molecules. This production of molecules can further be modified by other factors, such as sex hormones, infection, diabetes, and homocysteine levels. The inflammation-related molecules, which include adhesion molecules, mediate the transmigration of monocytes to the site of the lipid accumulation. The migrated monocytes proliferate and differentiate into macrophages. These macrophages, with the help of various receptors and ligands, take up the extensively modified lipid particles, leading to the formation of foam cells. As these cells die by apoptosis, the lipid-filled contents of them contribute to the necrotic core of the lesion. Smooth muscle cells (SMCs) further increase the size of the atherosclerotic plaque: they migrate to the site of the lesion, stimulated by cytokines and growth factors, which are expressed by macrophages. These intimal SMCs then contribute to the formation of a fibrous cap by secreting fibrous elements and extracellular matrix. Processes leading to the development of a fibrous lesion are further modified by risk factors such as hormones, elevated homocysteine, hypertension, and infection. (Lusis 2000)

Products of the inflammatory cells in atherosclerotic lesions participate in maintenance and degradation of the fibrous cap of the plaque. A rupture of the cap and subsequent thrombosis can lead to acute coronary and cerebrovascular events, and plaque composition and vulnerability are important determinants of future complications. Stability of an atherosclerotic lesion is influenced by several factors, including proteinases that degrade extracellular matrix, calcification of the lesion, factors contributing to inflammation, and also infection. A plaque rupture leads to a coagulation cascade, which results in the formation of a thrombus or a blood clot. Tissue factor, plasminogen activator, and other important molecules mediate these events. (Lusis 2000)

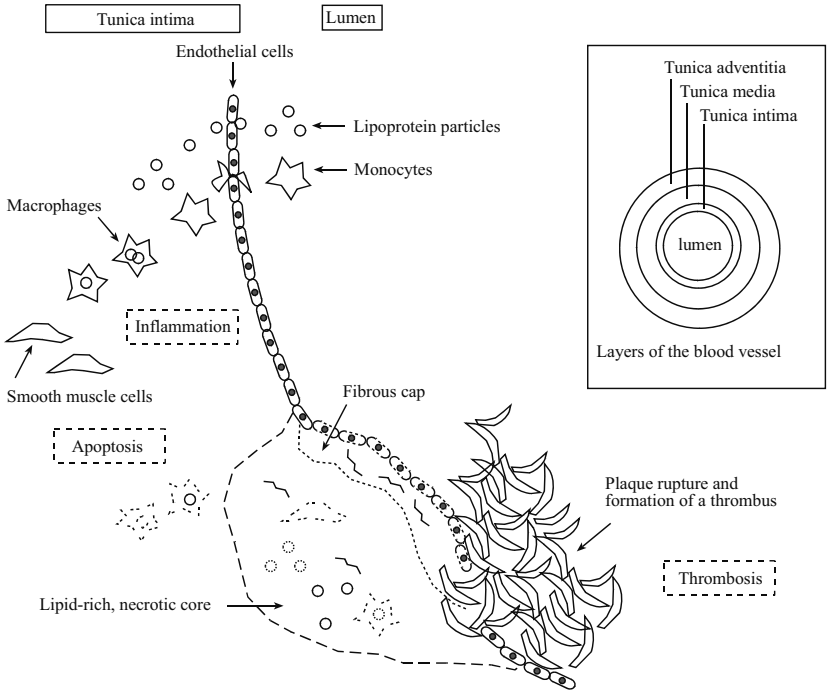

Figure 3. A simplified representation of the development of atherosclerotic plaques. Lipoprotein accumulation in the subendothelial matrix is followed by recruitment of monocytes and smooth muscle cells. Foam-cell formation, inflammation reactions, and apoptosis contribute to the formation of a lipid-rich, necrotic core of the lesion. Development of a fibrous cap and its subsequent rupture can lead to formation of a thrombus and possibly in ischemic cardiovascular events. Modified from Lusis 2000 and Watkins and Farrall 2006.

The atherosclerotic processes usually take decades to develop into a phase where a vessel narrowing or occlusion by thrombosis causes clinical complications. In ischemic heart disease, or coronary heart disease (CHD), atherosclerosis affects the arteries of the heart. Atherosclerosis can also affect arteries of the brain (cerebrovascular diseases), or the circulatory system outside of the brain and heart (peripheral vascular disease). The clinical complications caused by atherosclerosis are due to shortage of oxygen in the tissues whose oxygen is provided by the occluded blood vessels.

Narrowed coronary arteries fail to supply the tissue of the heart with an adequate amount of blood and oxygen. Angina pectoris, chest pain with variable severity, is a

common symptom of the restricted blood flow in the coronary arteries and can present both stable and unstable forms (Grech and Ramsdale 2003). The most severe outcome of coronary atherosclerosis is myocardial infarction (MI) (Achar et al. 2005), preceded by a rupture of an atherosclerotic plaque and a formation of a thrombus. In MI, tissue of the heart suffers severe damage due to interrupted blood flow and oxygen supply. MI is a life-threatening event, requiring immediate medical care.

Besides MI, another acute outcome of atherosclerosis of the arteries is ischemic stroke, where blood flow in the brain tissue is decreased or disrupted. Although ischemic stroke usually is a result of a local damage to the vessel wall from atherosclerosis, these strokes can also result from an embolus of cardiac origin. Cerebral infarction is a severe consequence of the impaired supply of oxygen and results in permanent cellular damage. The extent of the damage is dependent upon the duration of the ischemia and the availability of collateral circulation. (Frizzell 2005)

In peripheral arterial disease (PAD), blood flow is restricted in narrowed or occluded arteries leading to the stomach, arms, legs, feet, or kidneys. Intermittent claudication, a condition where muscle pain or cramping in legs or arms is triggered by activity, is a common symptom of the restricted circulation in the limbs (Leng et al. 1996).

Aortic atherosclerotic disease is a significant risk factor for embolic stroke and peripheral embolization, especially when there are severe atherosclerotic plaques in the aorta. Aortic plaques can also be seen as an expression of generalized atherosclerosis and they correlate with higher prevalence of CHD. (Kronzon and Tunick 2006)

2.3.1 Cardiovascular risk factors

Several environmental, biochemical, and genetic risk factors contribute to the risk of CVD (Table 4) (Yusuf et al. 2004; Goldstein et al. 2006).

Table 4. Risk factors for cardiovascular disease.

Risk factor	Environmental risk factor	Physiological or biochemical risk factor	Genetic risk factor
Alcohol consumption	X		X
Smoking	X		X
Lack of physical activity	X		X
Diet	X		X
Obesity, body composition		X	X
Diabetes and blood glucose levels		X	X
Hypertension		X	X
Low blood HDL levels		X	X
High blood LDL levels		X	X
High blood triglyceride levels		X	X
Thrombotic factors		X	X
Family history of CVD			X
Male gender			X
Age			(X)

Some of the CVD susceptibility is due to age, gender, and family history of CVD, and cannot be modified by lifestyle changes. Adjusting dietary habits, exercise, and cutting down on smoking and alcohol consumption, however, can reduce the risk of CVD (Yusuf et al. 2004; Goldstein et al. 2006). These lifestyle changes may reduce the risk of CVD through changes in some conventional physiological or biochemical risk factors, such as levels of serum lipoproteins, obesity, or hypertension.

Lipoproteins such as LDL have a significant role in the formation of atherosclerotic plaques, and clinical trials have shown the beneficial effect of lipid lowering medication on the risk of CVD. Other major components of a lesion and thrombus formation have also associated with CVD in epidemiological studies; these risk factors include elevated levels of inflammatory molecules such as C-reactive protein, haemostatic factors (e.g. fibrinogen and plasminogen activator inhibitor type 1) and homocysteine. (Lusis 2000)

Many of the established risk factors for CVD have a strong genetic component. Recent studies have also suggested that many of the risk factors conventionally considered as solely environmental (Table 4), might have a genetic predisposition (Whitfield et al. 2004; Ingelsson et al. 2007; Keskitalo et al. 2007; Uhl et al. 2007). Suggestive evidence for the role of infectious agents in atherosclerosis also exist (Fong 2002; Mussa et al. 2006).

Individuals with common metabolic disorders, such as familial combined hyperlipidemia (FCHL) (Goldstein et al. 1973a) and the metabolic syndrome (MetS) (Reaven 1993; Eckel et al. 2005), are also at increased risk of CVD (Austin et al. 2000; Voors-Pette and de Bruin 2001; Lakka et al. 2002; Hopkins et al. 2003; Malik et al. 2004).

2.3.2 Phenotypic overlap between cardiovascular disease, metabolic syndrome and familial combined hyperlipidemia

Extensive phenotypic overlap exists between MetS (Reaven 1993; Alberti et al. 2005; Eckel et al. 2005), FCHL (Goldstein et al. 1973a), and type 2 diabetes mellitus (T2DM) (Alberti and Zimmet 1998), which all predispose to premature CVD (Tuomilehto et al. 1996; Austin et al. 2000; Voors-Pette and de Bruin 2001; Lakka et al. 2002; Hopkins et al. 2003; Malik et al. 2004; Huxley et al. 2006). T2DM primarily involves insulin resistance and hyperglycemia, but also often associates with hypertension, obesity, and dyslipidemias (Alberti and Zimmet 1998). FCHL, first described in 1973, is a dyslipidemia characterized by elevated serum total cholesterol, triglycerides, or both (Goldstein et al. 1973a). Diagnostic criteria of FCHL also include at least one affected first degree relative and usually also premature CHD of the proband.

MetS, a common metabolic disorder, is characterized by a clustering of CVD risk factors (Alberti et al. 2005; Eckel et al. 2005). The prevalence of MetS, which is highly age-dependent, has greatly increased during the last few decades alongside with the increasing rates of obesity and T2DM. Although MetS is more common in the older age-groups, the disorder and biomarkers of an increased risk of CVD can be found among obese children and adolescents.

Several definitions for MetS exist, proposed by the World Health Organization (WHO), National Cholesterol Education Program's Adult Treatment Panel III (NCEP:ATP III), the European Group for the Study of Insulin Resistance (EGIR), and the International Diabetes Federation (IDF). These definitions differ in their cut-off point values for the components of MetS, and may also have their own ways of combining the components to define MetS. Most of them do, however, agree on a

central set of criteria, which include impaired glucose tolerance, impaired fasting glucose or insulin resistance, dyslipidemia, central obesity, and hypertension. Overabundance of free fatty acids plays a central role in the pathophysiology of MetS. Along with other factors, it contributes to several unfavorable outcomes; increased production of glucose by the liver, reduction in HDL cholesterol and increase in small dense (more atherogenic) LDL, hypertension, and increases in proinflammatory cytokines. (Eckel et al. 2005)

Many of the components of MetS are established risk factors for CVD, and they also include features associated with T2DM (impaired glucose tolerance and impaired fasting glucose). This overlap between T2DM, FCHL, and MetS suggests that the same genetic determinants may contribute to the pathology.

2.3.3 Gender-specific aspects of cardiovascular disease

In the past, cardiovascular disease in men gained most of the attention and research was often conducted using study samples consisting of only men (Barrett-Connor 1997). As previously, in the younger age-groups the majority of patients whom die due to CVD today are still men (Tunstall-Pedoe 1998; Tilastokeskus 2007a). Women, however, outnumber men in CVD deaths in the older age groups (Figure 4) (Tunstall-Pedoe 1998; Tilastokeskus 2007a). In Finland, cardiovascular disease is correspondingly shifting from being a disease of middle-aged men to a disease of elderly women (Kattainen et al. 2006).

Women undergo CVD at a later age compared to men (Figure 4). In 2006, 20% of men who died from CVD were under 65, 42% were under 75, and 79% were under 85 (Tilastokeskus 2007a). Among women, the corresponding figures were 5%, 13%, and 48%. Mortality of women began to increase at the age of 65 and the number of deaths was highest among 85 year olds, whereas among men the number of CVD deaths began to rise at the age of 50 and was highest at the age of 75 (Figure 4). Importantly, in all age-groups the proportion of men with CVD death out of all men at risk of death was higher than the proportion of women with CVD death out of all women at risk of death (Figure 5) (Tilastokeskus 2007b). As morbidity from CVD strongly correlates with mortality from it (Tunstall-Pedoe et al. 1994); the risk of incident CVD and related death is higher among men than among women, throughout life (Figure 5) (Tunstall-Pedoe 1998).

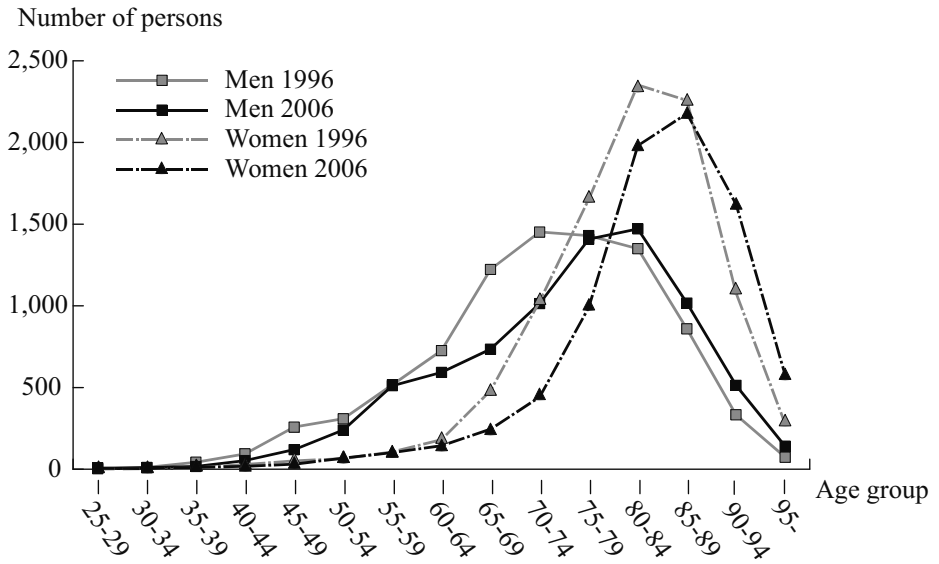

Figure 4. Number of deaths due to ischemic heart disease or cerebrovascular disease in Finland by sex and age, 1996 and 2006. (Tilastokeskus 2007a)

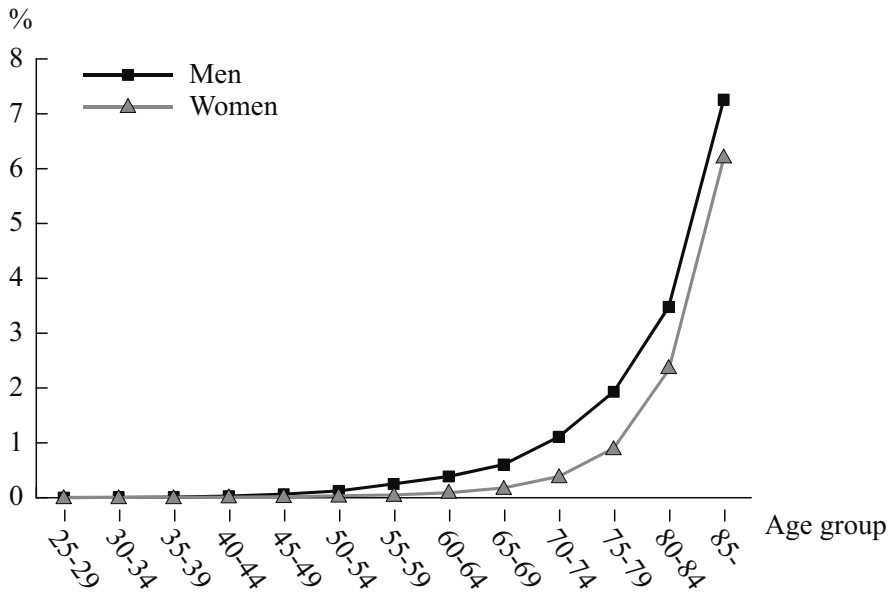

Figure 5. Mortality (% of population) due to ischemic heart disease or cerebrovascular disease in Finland by sex and age group, year 2006. (Tilastokeskus 2007a; Tilastokeskus 2007b)

Distribution of cardiovascular risk factors is often different among men and women (Barrett-Connor 1997; Weiss et al. 2006), and usually men possess the more unfavorable risk profile (Barrett-Connor 1997). Variation in environmental risk factors is, however, insufficient to fully explain the contrast of the frequency of CVD between the two sexes (Barrett-Connor 1997). Some differences are likely to be explained by the contribution of tissue-specific hormonal environments and gene expressions, which are dissimilar in men and women (Rinn and Snyder 2005). Sex hormones and their receptors play an important part in the etiology of CVD, participating in the activation of CVD-related gene expression in the smooth muscle cells, arterial endothelium, and liver (Barrett-Connor 1997; Mendelsohn and Karas 2005; Rinn and Snyder 2005).

2.4 Genetics of cardiovascular disease

The majority of CVD events result from a complex interplay between several genes and environmental risk factors (Figure 6), although rare monogenic forms of cardiovascular disease also exist.

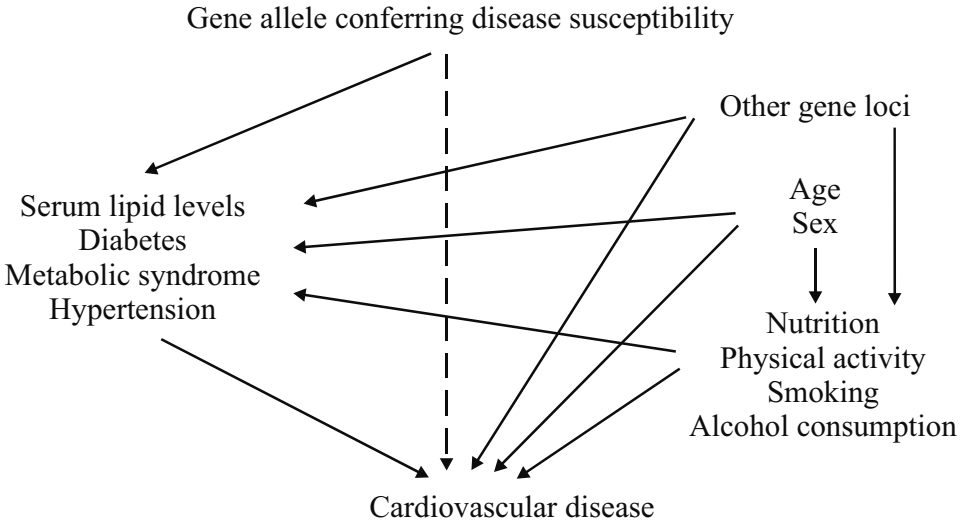

Figure 6. CVD is a complex disease. Association between a particular genetic risk factor and the disease can be in part due to intermediate phenotypes increasing susceptibility. Detection of the association between a genetic risk factor and the disease (depicted with a dashed line) is obscured by other genetic factors, and environmental factors contributing to the disease.

A large portion of the risk for CVD is considered to be explained by genetic factors (Zdravkovic et al. 2002; Flossmann et al. 2004; Fischer et al. 2005; Moskau et al. 2005; Wienke et al. 2005) (Table 5). For relatives of a CHD patient, the risk of CHD can be twofold (Jousilahti et al. 1996; Lloyd-Jones et al. 2004) and the risk of a stroke is also increased (Kiely et al. 1993; Schulz et al. 2004). A proportion of the increased CVD risk may be explained by genetic factors that influence intermediate CVD phenotypes, such as LDL or HDL levels (Table 5).

Table 5. Heritability estimates of stroke, CHD, and their risk factors

Trait	Heritability	Reference
T2DM / glucose / insulin	0.24-0.85	(Poulsen et al. 1999; Hanson et al. 2001; McQueen et al. 2003; Mills et al. 2004; Souren et al. 2007)
MetS	0.38-0.61	(McQueen et al. 2003; Bayoumi et al. 2007)
Obesity/BMI	0.31-0.94	(O'Connell et al. 1988; Brown et al. 2003; McQueen et al. 2003; Mills et al. 2004; Souren et al. 2007)
Hypertension / blood pressure	0.17-0.49	(Brown et al. 2003; McQueen et al. 2003; Mills et al. 2004)
Total cholesterol	0.32-0.75	(O'Connell et al. 1988; Heller et al. 1993; Mills et al. 2004; Souren et al. 2007)
Triglycerides	0.28-0.72	(Heller et al. 1993; McQueen et al. 2003; Mills et al. 2004)
LDL cholesterol	0.43-0.78	(O'Connell et al. 1988; Mills et al. 2004; Souren et al. 2007)
HDL cholesterol	0.42-0.79	(O'Connell et al. 1988; Heller et al. 1993; McQueen et al. 2003; Mills et al. 2004; Souren et al. 2007)
apoB	0.51-0.78	(Heller et al. 1993)
Coronary heart disease / atherosclerosis	0.05-0.84	(Zdravkovic et al. 2002; Fischer et al. 2005; Wienke et al. 2005; Cassidy-Bushrow et al. 2007)
Stroke	0.17-0.32	(Bak et al. 2002)

Several linkage studies have scanned the human genome in order to find chromosomal regions contributing to CVD (Pajukanta et al. 2000; Francke et al. 2001; Broeckel et al. 2002; Harrap et al. 2002; Chiodini and Lewis 2003; Hauser et al. 2004; Wang et al. 2004). Replication of the identified loci has been difficult and pinpointing the causative gene has often been challenging (Chiodini and Lewis 2003; Wang et al. 2004). Part of this challenge is explained by limitations of the linkage approach: When several loci contribute to the disease susceptibility and those with modest effects are hard to detect by linkage, background noise may cause different selection of loci to be identified in different studies (Chiodini and Lewis 2003; Watkins and Farrall 2006). Recently, genome-wide association scans

searching for gene loci contributing to CVD have also become feasible (Larson et al. 2007; O'Donnell et al. 2007).

2.4.1 Monogenic and multifactorial forms of cardiovascular disease and CVD related traits

Many studies have had success in finding genetic loci contributing to specific risk factors for CVD, instead of the complex phenotype for CVD (Watkins and Farrall 2006). Several lipid metabolism related disorders are monogenic diseases that increase the risk for CVD substantially (Nabel 2003; Watkins and Farrall 2006). These include familial hypercholesterolemia (mutated gene *LDLR*) (Goldstein and Brown 1979), familial defective apolipoprotein B-100 (mutated gene *APOB*) (Innerarity et al. 1987; Soria et al. 1989), and sitosterolemia (mutated genes *ABCG5* and *ABCG8*) (Berge et al. 2000; Lee et al. 2001) (Table 6). One person out of five hundred carries the LDLR gene mutation, and only one in a million is a homozygote, who develops severe atherosclerosis and suffers a myocardial infarction early in life (Nabel 2003; Watkins and Farrall 2006). Patients affected with sitosterolemia are even fewer, less than 100 have been reported in literature worldwide (Lee et al. 2001).

Another suggested risk factor for myocardial infarction, stroke, or especially venous thromboembolic disease is factor V Leiden mutation, which prevents the degradation of factor V and enhances clot formation (Bertina et al. 1994; Nabel 2003). The frequency of factor V Leiden mutation in northern Europeans is 2-7%. Risk of thromboembolic events is also increased in homocystinuria, a disorder with an incidence of approximately 1 in 335,000 (Naughten et al. 1998).

Monogenic diseases that elevate or lower blood pressure also exist (Table 6).

Studies on these monogenic disorders have identified targets for developing for instance treatments for high serum lipids. They have also provided valuable information on the biological pathways and mechanisms of diseases for the common, complex forms of CVD.

Table 6. Examples of genes associated with CVD or related traits with Mendelian inheritance

Syndrome	Mechanism	Reference
Hypercholesterolemia, autosomal dominant (Familial hypercholesterolemia)	Nonfunctional receptor of LDLR fails to take up plasma cholesterol	(Goldstein and Brown 1979)
Hypercholesterolemia, autosomal dominant, type b (apolipoprotein b-100, familial ligand-defective)	Clearance of plasma LDL is decelerated as binding of apoB-100 to LDLR receptors is reduced due to mutations in the *APOB* gene	(Innerarity et al. 1987; Soria et al. 1989)
Sitosterolemia	Mutated forms of *ABCG5* and *ABCG8* genes predispose to sterol overaccumulation	(Berge et al. 2000; Lee et al. 2001)
Hypercholesterolemia, autosomal dominant	Missense mutations, probably gain-of-function mutations, in the *PCSK9* gene cause hypercholesterolemia	(Abifadel et al. 2003)
Tangier disease	Accumulation of cellular cholesterol and low plasma HDLc levels due to mutations in the *ABCA1* gene	(Bodzioch et al. 1999; Brooks-Wilson et al. 1999; Rust et al. 1999)
Homocystinuria	Increased risk for thromboembolic events due to mutated *CβS* gene	(Naughten et al. 1998; Yap 2003)
Thrombophilia due to deficiency of cofactor for activated protein c, Leiden type	Mutated *F5* gene prevents the degradation of factor V and enhances clot formation	(Bertina et al. 1994)
Cortisol 11-beta-ketoreductase deficiency (Apparent mineralocorticoid excess)	Mutations in the *HSD11B2* gene lead to increased blood pressure	(White et al. 1997)
Glucocorticoid-remediable aldosteronism	*CYP11B2* gene is fused with the *CYP11B1* gene, which leads to changes in control of gene expression and thus increased blood pressure	(Lifton et al. 1992)
CAD, autosomal dominant, 1 (ADCAD1)	Deletion in exon 11 of the *MEF2A* gene result in loss of function of transcriptional activation activity	(Wang et al. 2003) (Later argued against by (Weng et al. 2005))

Common forms of CVD are characterized by the interplay of several environmental risk factors and many genes with possibly multiple different alleles contributing to disease susceptibility (Lohmueller et al. 2003). Interesting candidate genes include signal molecules from inflammation pathways, apolipoproteins and other well known lipid metabolism related genes (Figure 3) (Casas et al. 2006; Watkins and Farrall 2006). Since many of the genes and pathways contributing to the progression and emerging of CVD are related (for example the impact of lipids and inflammation on

the atherosclerotic lesion), especially transcription factors regulating the expression of several genes are plausible candidate genes for CVD studies.

Well-known genes implicated in the etiology of the common CVD include apolipoprotein genes, such as apolipoprotein E (*APOE*) and *APOB* genes, thrombosis and hemostasis related genes, such as fibrinogen, factor V, and prothrombin, and inflammation related genes, such as the interleukin-6 (*IL6*) gene (Table 7). Many genes have also been suggested in the etiology of T2DM, and some gene associations, such as peroxisome proliferator-activated receptor-γ (*PPARG*) and transcription factor 7-like 2 (*TCF7L2*), have been confirmed in multiple samples. New candidate genes for CVD and its risk factors are located with traditional methods and with emerging large-scale genome-wide association scans (Helgadottir et al. 2007; Larson et al. 2007; Levy et al. 2007; McPherson et al. 2007; Samani et al. 2007). Candidate genes are further studied in large international collaborations (Evans et al. 2005; Danesh et al. 2007).

Table 7. Examples of genes suggested in the genetics of common forms of CVD or its risk factors

Gene	Cardiovascular outcome/risk factor	Reference
ABCA1	Dyslipidemia, CHD	(Tregouet et al. 2004)
ACE	CHD, MI	(Morgan et al. 2003)
AGT	Hypertension	(Sethi et al. 2003)
ALOX5AP	MI and stroke	(Helgadottir et al. 2004)
APOE	Dyslipidemia, CHD, MI	(Song et al. 2004)
APOA5	Dyslipidemia	(Martin et al. 2003b)
APOB	Dyslipidemia, CHD, MI	(Chiodini et al. 2003)
CETP	Dyslipidemia, CHD	(Boekholdt et al. 2005)
F2 (Prothrombin)	MI	(Burzotta et al. 2004)
F5	Venous thromboembolic events, MI, stroke	(Kim and Becker 2003)
Fibrinogen	MI	(Boekholdt et al. 2001)
FTO	Obesity	(Frayling et al. 2007)
IL6	CHD, MI	(Georges et al. 2001; Tanaka et al. 2005)
INSIG2	Obesity	(Herbert et al. 2006)
MTHFR	CHD	(Klerk et al. 2002)
PDE4D	Stroke	(Gretarsdottir et al. 2003)
PPARG	TD2M	(Altshuler et al. 2000)
SERPINE1 (PAI-1)	MI	(Boekholdt et al. 2001)
TCF7L2	TD2M	(Grant et al. 2006)

Modified from Casas et al. 2006, Arnett et al. 2007, and Cambien and Tiret 2007.

2.4.2 Sex-specific differences in the genetics of cardiovascular disease

As well as in sex-specific gene expression, the two sexes can also differ in regards to the heritability of CVD risk factors, such as LDL, HDL, and systolic blood pressure (Weiss et al. 2006). Correspondingly, studies on CVD have observed sex-specific results (Tentschert et al. 2003; Weiss et al. 2006). Sometimes these sex-dependent results may be due to complicating environmental factors which prevent detection of a genetic association in one of the sexes despite an existing association. In some cases a risk increasing effect of a risk factor may be larger in one of the sexes; For instance low HDL levels or diabetes increase the risk of CVD especially among women (Barrett-Connor 1997; Dale et al. 2007). Hormone-replacement therapy at menopause does not necessarily decrease the risk of CVD among women (Waters et al. 2004), although it has been suggested that the onset of the therapy or the physiological differences between the arteries of women in different age groups have not sufficiently been accounted for in previous studies (Mendelsohn and Karas 2005).

2.4.3 *USF1* as a candidate gene for cardiovascular disease

The disorders FCHL, metabolic syndrome, and type 2 diabetes are often implicated in cardiovascular disease (Genest et al. 1992; Tuomilehto et al. 1996; Austin et al. 2000; Voors-Pette and de Bruin 2001; Lakka et al. 2002; Hopkins et al. 2003; Malik et al. 2004; Huxley et al. 2006). A chromosomal region at 1q21-23 has showed compelling evidence of linkage to these three disorders (Hanson et al. 1998; Pajukanta et al. 1998; Elbein et al. 1999; Vionnet et al. 2000; Hsueh et al. 2003; Ng et al. 2004a; Ng et al. 2004b; Xiang et al. 2004), and subsequently an interesting candidate gene within the region has been identified (Pajukanta et al. 2004). This gene encodes upstream transcription factor 1 (USF1).

USF1 is a ubiquitously expressed transcription factor regulating transcription of several genes from the lipid and glucose metabolism pathways (Naukkarinen et al. 2005) (Table 8). The genomic size of the gene encoding USF1 is 6.7 kb, it has two major splicing variants (Saito et al. 2003), and resides in chromosome 1q22-23 (Figure 7 and Figure 8) (Shieh et al. 1993).

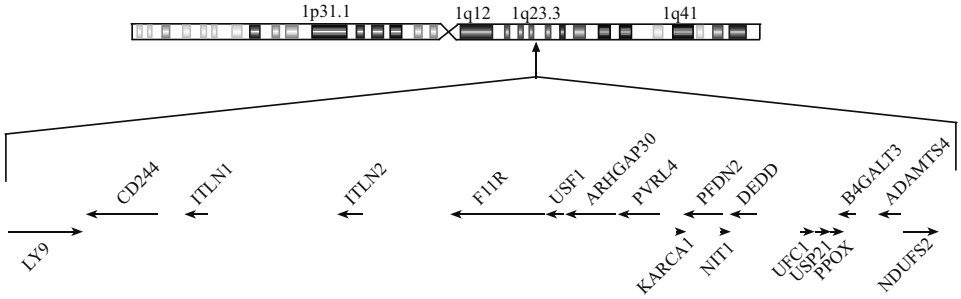

Figure 7. The chromosomal region containing *USF1*. Modified from the UCSC Genome Browser.

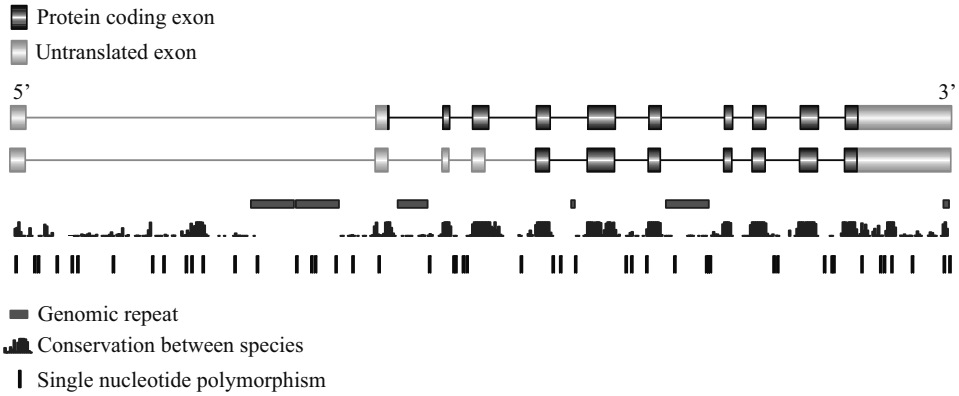

Figure 8. The *USF1* gene. The two common splicing forms of *USF1* contain 11 exons, from which seven or ten are translated into a protein sequence. Exon 4 of the gene contains an alternative splicing site and the exon is thus 31 bp shorter in the alternative splicing variant. Modified from the UCSC Genome Browser and Saito et al. 2003.

The USF1 protein, which comprises 310 amino acids and contains a helix-loop-helix motif, binds an E-Box motif (5'-CACGTG-3') in the promoter regions of its target genes as a dimer with another USF1 or with a USF2 protein (Sawadogo 1988; Sawadogo et al. 1988; Gregor et al. 1990; Roy et al. 1991; Ferre-D'Amare et al. 1994; Viollet et al. 1996). Besides USF2, USF1 possibly interacts with other proteins such as GTF2I, FOSL1, ESR1, and NFYA (Peri et al. 2003; Mishra et al.

2006; Stark et al. 2006). The Cha-protein has been indicated to regulate the activity of USF1 (Rodriguez et al. 2003).

One of the cardiovascular-related genes regulated by USF1 is angiotensinogen (*AGT*), which participates in blood pressure regulation (Caulfield et al. 1994; Saito et al. 2003). In addition USF1 affects the expression of some important genes involved in cholesterol homeostasis and fatty acid synthesis, such as *APOE* and fatty acid synthase (*FASN*) (Casado et al. 1999; Salero et al. 2003; Naukkarinen et al. 2005). Furthermore, USF1 has a role in the regulation of some inflammation related genes: hepcidin (Bayele et al. 2006), beta-2 microglobulin (Gobin et al. 2003), and metallothionein (Andrews 2000). See Table 8.

The function of USF1 as a regulator of these and other important genes suggests *USF1* is an important candidate gene for CVD.

Table 8. USF1 is indicated in the regulation of several genes relevant to cardiovascular disease

Gene		Biological processes by Gene Ontology[a]
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	cholesterol efflux, interleukin-1 beta secretion, intracellular cholesterol transport, platelet dense granule organization and biogenesis, reverse cholesterol transport
ACACA	acetyl-Coenzyme A carboxylase alpha	acetyl-CoA carboxylase activity
ACTC1	actin, alpha, cardiac muscle 1	muscle contraction, regulation of heart contraction
AGT	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	cell surface receptor linked signal transduction, cell-cell signaling, regulation of blood pressure
APOA2	apolipoprotein A-II	glucose metabolic process, lipid transport, regulation of lipid catabolic and metabolic process, neutrophil activation, positive regulation of interleukin-8 biosynthetic process, regulation of cytokine production, response to glucose stimulus
APOA5	apolipoprotein A-V	lipid transport, regulation of lipid biosynthetic process, tissue regeneration, triacylglycerol metabolic process
APOC3	apolipoprotein C-III	triacylglycerol metabolic process
APOE	apolipoprotein E	blood circulation, cholesterol homeostasis, induction of apoptosis, lipid transport, lipoprotein metabolic process, response to reactive oxygen species, cholinergic
APP	amyloid beta (A4) precursor protein	cellular copper ion homeostasis, neuromuscular process
B2M	beta-2-microglobulin	immune response
BDNF	brain-derived neurotrophic factor	nervous system development
BRCA2	breast cancer 2, early onset	double-strand break repair, nucleotide-excision repair, regulation of S phase of mitotic cell cycle, regulation of response to estrogen stimulus
CCNB1	cyclin B1	G2/M transition of mitotic cell cycle
CCND1	cyclin D1	G1/S transition of mitotic cell cycle, positive regulation of cyclin-dependent protein kinase activity, positive regulation of protein amino acid phosphorylation
CDK4	cyclin-dependent kinase 4	G1/S transition of mitotic cell cycle, positive regulation of fibroblast proliferation, regulation of gene expression
CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	angiogenesis, cell migration, homophilic cell adhesion, integrin-mediated signaling pathway
CEACAM5	carcinoembryonic antigen-related cell adhesion molecule 5	integral to plasma membrane
CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	cell-cell signaling, signal transduction
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha	cytokine and chemokine mediated signaling pathway, generation of precursor metabolites and energy
CXCR4	chemokine (C-X-C motif) receptor 4	apoptosis, chemotaxis, elevation of cytosolic calcium ion concentration, immune

		response, inflammatory response, response to virus
CYP19	cytochrome P450, family 19, subfamily A, polypeptide 1	steroid biosynthetic process
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	oxygen binding
CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	lipid metabolic process
FASN	fatty acid synthase	protein binding
FSHR	follicle stimulating hormone receptor	female gamete generation, female gonad development, G-protein coupled receptor protein signaling pathway, spermatogenesis
GCGR	glucagon receptor	generation of precursor metabolites and energy, regulation of blood pressure, response to nutrient
GHRL	ghrelin/obestatin preprohormone	adult feeding behavior, cortisol secretion, elevation of cytosolic calcium ion concentration, glucose metabolic process, growth hormone secretion, hormone-mediated signaling, negative regulation of angiogenesis, negative regulation of apoptosis, negative regulation of endothelial cell proliferation, negative regulation of inflammatory response, negative regulation of interleukin-1 beta production, negative regulation of interleukin-6 biosynthetic process, negative regulation of tumor necrosis factor biosynthetic process, positive regulation of appetite, positive regulation of insulin secretion, response to estrogen stimulus
HBB	hemoglobin, beta	nitric oxide transport, oxygen transport, positive regulation of nitric oxide biosynthetic process
HAMP	hepcidin antimicrobial peptide	immune response
HOXB4	homeobox B4	multicellular organismal development, regulation of transcription, DNA-dependent
LIPE	lipase, hormone-sensitive	fatty acid metabolic process, generation of precursor metabolites and energy, protein amino acid phosphorylation, response to drug
MAP2K1	mitogen-activated protein kinase 1	protein binding
PF4	platelet factor 4 (chemokine (C-X-C motif) ligand 4)	leukocyte chemotaxis, negative regulation of angiogenesis, negative regulation of megakaryocyte differentiation, platelet activation
PIGR	polymeric immunoglobulin receptor	protein secretion
PTPN6	protein tyrosine phosphatase, non-receptor type 6	apoptosis, G-protein coupled receptor protein signaling pathway, protein amino acid dephosphorylation
REN	renin	angiotensin maturation, proteolysis, regulation of blood pressure
SERPINE1	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	blood coagulation, fibrinolysis
SPP1	secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)	cell-matrix adhesion, negative regulation of bone mineralization, ossification
TERT	telomerase reverse transcriptase	telomere maintenance

[a] if more than two processes were described by Gene Ontology, a few relevant to lipid metabolism, glucose metabolism, or inflammation were selected for this table. Modified from Ashburner et al. 2000, Matys et al. 2003, and Naukkarinen et al. 2005.

3 AIMS OF THE STUDY

When this study was initiated, the *USF1* gene on chromosome 1q21-23 had been associated with familial combined hyperlipidemia in Finnish FCHL families (Pajukanta et al. 2004). *USF1* was also known to regulate several genes involved in pathways relevant for cardiovascular disease. We thus sought to investigate the relation between allelic variation within *USF1* and cardiovascular disease and its risk factors. The specific aims were, in our population-based study samples:

1) To study a potential relationship between *USF1* alleles and traditional cardiovascular risk factors such as serum lipid values, the metabolic syndrome, and body composition measurements (I, III).
2) To further determine whether *USF1* alleles affect atherosclerotic phenotypes at the target tissue level, i.e. the vessel wall (II).
3) To examine if the allelic variants of *USF1* affect the risk of a sudden cardiac death or a cardiovascular event at the population level (I-III).
4) To investigate how the use of whole genome amplified DNA affects accuracy of genotyping when some of the study samples have a low yield of DNA (IV).

4 MATERIALS AND METHODS

4.1 Table of materials and methods

The details of the materials and methods used in this study are described in the original publications (I-IV) according to Table 9.

Table 9. Materials and methods used in this study.

Material or method	Original publication
Study samples	
ULSAM	I
HSDS	II
FINRISK	III
Whole genome amplified FINRISK samples	IV
Control DNA	I, IV
Definition of metabolic syndrome	I
Definition of cardiovascular disease	I - III
Definition and measurements of atherosclerosis	II
Measurements of lipid parameters	I, III
Laboratory procedures	
DNA extraction	I - IV
DNA quantitation	IV
Polymerase chain reaction (PCR)	I - IV
Testing of DNA integrity	IV
Agarose gel electrophoresis	I - IV
Sequencing	IV
Allele-specific primer extension on microarrays	II - IV
Sequenom MALDI-TOF mass spectrometry	II - IV
Single-base primer extension assay with fluorescence polarization detection	I
Multiply-primed rolling circle DNA amplification	III, IV
Statistical methods	
SNP selection	I - IV
LD estimation	I - III
Haplotype estimation	I - III
Chi-squared tests	I - III
Analysis of covariance	I, III
Regression analysis	II
Time-to-event analysis	I, III
Mixed longitudinal analysis	I
Permutation of association	I, III
False discovery rate correction	II

Table 9 continued.

Material or method	Original publication
Analysis programs	
SNPSnapper	II - IV
Haploview	I - III
PHASE 2.11	I, III
SpectroDESIGNER	II - IV
SpectroCALLER	II - IV
Primer3	II - IV
Oligo Analyzer 3.0	II - IV
QuantArray 3.0	II - IV
GeneMapper 3.0	IV
Sequencher 4.1.4	IV
SAS for Windows (v8.2 & 9.1)	I-III

4.2 Study samples

Three different population-based study samples were included in this study (Table 10).

Table 10. Characteristics of the study samples.

Characteristic	ULSAM	HSDS	FINRISK
Recruitment areas	Uppsala county, Sweden	Helsinki area, Finland	5 regions of Finland
Genotyped subjects (n)	2,322	700	2,225
Age distribution at the baseline of the study (years)	50	33-70	25-74
CVD cases (n)	701	230	1,190
Subjects free of CVD (n)	998	470	626 (the healthy subcohort)
Proportion of men in the sample	100%	100%	69%
Length of follow-up period (years)	32	no follow-up	7 and 10

The Uppsala Longitudinal Study of Adult Men (ULSAM) was approved by the Ethics Committee of the Medical Faculty of Uppsala University, and the FINRISK studies were approved by the ethical committee of the National Public Health Institute. All ULSAM and FINRISK participants gave informed consent and the principles of the Helsinki Declaration were followed. The Ethics Committee of the Department of Forensic Medicine of the University of Helsinki approved the Helsinki Sudden Death Study (HSDS). Samples for the study of whole genome amplification were selected from the FINRISK study sample.

4.2.1 The ULSAM study sample (I)

The ULSAM study sample consisted of 2,322 men born between 1920-1924 and living in Uppsala, Sweden, and recruited at the age of 50 (Table 10). The participants were followed for 32 years (1970-2002) with medical interviews and measurements at the ages of 50, 60, 70, and 77 (Table 11). Information on the cardiovascular events of the participants was obtained from the National Hospital Discharge Register and the National Causes of Death Register.

Table 11. The ULSAM follow-up cohort.

	Survey I	**Survey II**	**Survey III**	**Survey IV**	**End of follow-up**
Age	50 years	60 years	70 years	77 years	
Year of survey	1970-1973	1981-1984	1991-1995	1997	2002
Number of participants examined	2,322	1,860	1,221	839	
CVD events [a]	14	134	373	518	701
MetS ATPIII [b]	452	662	878	969	-
Total mortality	-	98	442	748	1,078
Ineligible	-	94	219	176	-
Measurements	Questionnaire Cholesterols [c] Blood pressure Fasting glucose BMI ApoA-1, ApoB-100 IVGTT	Questionnaire Cholesterols [c] Blood pressure Fasting glucose BMI, waist grid IVGTT	Questionnaire Cholesterols [c] Blood pressure Fasting glucose BMI, waist grid ApoA-1, ApoB-100 Insulin clamp Blood samples for DNA	Questionnaire Cholesterols [c] Blood pressure Fasting glucose Waist grid	

Numbers for CVD events, metabolic syndrome (MetS), and mortality are cumulative from the baseline.

[a] A coronary event or ischemic stroke during follow-up. Nine people with baseline CVD (50 years of age) had another event during follow-up.

[b] ATP III criteria (male-specific) used. Individuals with missing values excluded at each time point.

[c] Cholesterols: Total cholesterol, high-density lipoprotein cholesterol, triglycerides

4.2.2 The Helsinki Sudden Death Study sample (II)

Finnish middle-aged men who died suddenly outside of hospital in the Helsinki City area during 1981-1982 and 1991-1992 were included in the HSDS study (Table 10). Areas of different types of atherosclerotic lesions and coronary narrowing were measured in medicolegal autopsy of the study subjects. These measurements defined fatty streaks (areas stained red with Sudan IV and unaffected with other types of changes), fibrotic lesions (elevated plaque areas that exhibited no ulceration or thrombosis), complicated lesions (plaque areas with ulceration or thrombosis), and advanced lesion areas (formed by combining elevated plaque areas and complicated lesion areas). Calcification of the arteries was measured by x-rays.

4.2.3 The FINRISK study samples (III)

From the original FINRISK study samples, a case-cohort subset was included in this study of the *USF1* gene (Table 10, Figure 9, and Figure 10). In 1992 the original FINRISK -92 study invited 4,000 men and 4,000 women from 4 regions of Finland (Figure 9 and Figure 10) to a physical baseline examination. FINRISK -97 study invited 6,000 men and 5,500 women from 5 regions 5 years later (Figure 9 and Figure 10).

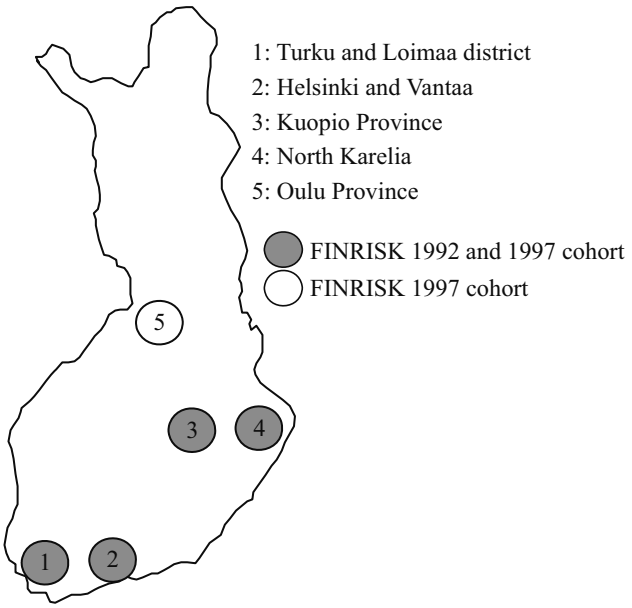

Figure 9. FINRISK study areas.

After the follow-up periods (FINRISK -92 ten years, FINRISK -97 seven years) participants with incident coronary heart disease or stroke during the follow-up, and participants with a venous thromboembolic event or death during follow-up, were selected for genotyping in this study (Figure 10). A random subcohort of the original follow-up samples was also genotyped (Figure 10). In addition, participants with CVD at the baseline of the study were selected for genotyping. Information on the events or death during the follow-up periods was obtained from specific myocardial infarction and stroke registers (FINAMI and FINSTROKE) and from the National Hospital Discharge Register and the National Causes of Death Register.

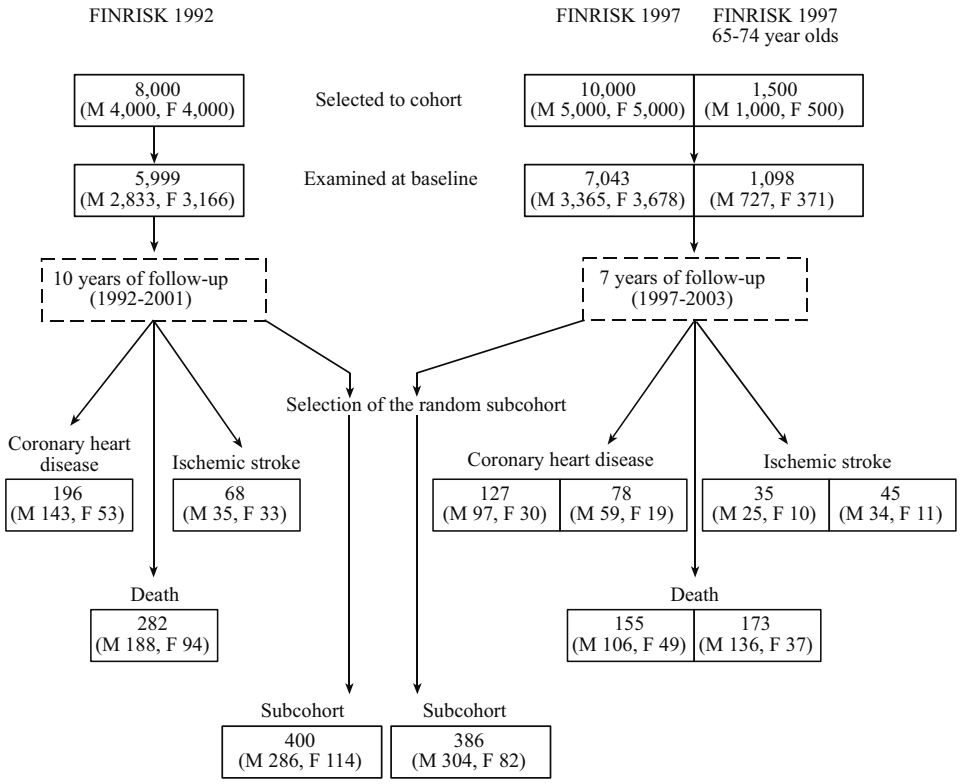

Figure 10. The FINRISK follow-up cohort. Number of participants at different stages of the follow-up. At the end of the follow-up, cases and random subcohorts were selected for genotyping and analyses. M=males, F=females.

4.2.4 Whole genome amplified FINRISK samples (IV)

FINRISK -92 DNA samples with an available DNA amount less than or equal to 7.5 µg were selected for the WGA study. A control DNA sample set comprised 36 high-yield samples that are regularly used as laboratory controls in our laboratory.

4.3 Laboratory methods

4.3.1 DNA extraction

DNA of ULSAM (I) participants was isolated with a standard salting-out method from EDTA blood samples drawn at 70 years. In addition, for some samples DNA was extracted from paraffin-embedded tissues with K digestion and QIAamp minikits (Qiagen Inc., CA, USA). In the HSDS (II) 1981-1982 series, DNA was extracted from paraffin-embedded samples of cardiac muscle with a modification of the method by Isola and colleagues (Isola et al. 1994). In the HSDS (II) 1991-1992 series, a standard phenol-chloroform method (Vandenplas et al. 1984) was used to extract DNA from frozen (-70°C) cardiac samples. The DNA of each FINRISK (III, IV) participant was extracted from 10 ml of EDTA-treated whole-blood samples using a standard phenol-chloroform method (Vandenplas et al. 1984) or salt precipitation.

4.3.2 Marker selection and genotyping

Genotyped *USF1* polymorphisms with their minor allele frequencies are shown in Figure 11.

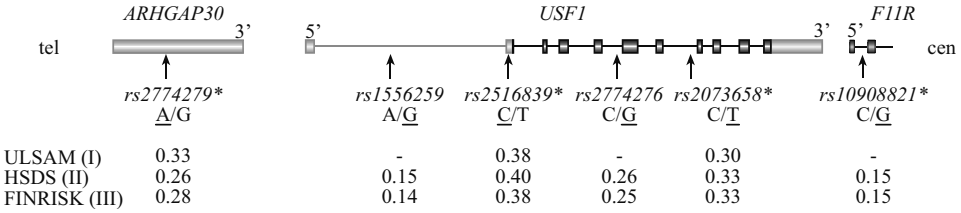

Figure 11. Minor allele frequencies of the genotyped *USF1* SNPs. Minor allele frequencies are for individuals free of CVD. Only the SNPs tagging the three most common *USF1* haplotypes were genotyped in the ULSAM (I) study (*rs2774279*, *rs2516839*, *rs2073658*). SNPs genotyped in the WGA study (IV) are indicated with an asterisk. The minor allele of each SNP is underlined. Genes *ARHGAP30* and *F11R* are only partially shown in the figure, and the distance between the genes is out of proportion. The distance from *rs2774279* to *rs10908821* is 9.02 kb. Tel=telomere, cen=centromere.

SNP selection was based on the haplotype structure of the *USF1* region and all common (frequency >5%) LD bins listed in the SeattleSNPs Variation Discovery Resource were covered with the six selected SNPs.

The genotyping methods included a single-base primer extension assay with fluorescence polarization detection (I), Sequenom's MALDI-TOF mass spectrometry (II-IV), and allele-specific primer extension on microarrays (II-IV).

4.3.3 Whole genome amplification

FINRISK samples with a low DNA yield were amplified for the whole genome with GenomePhi DNA amplification kit (Amersham Biosciences, Piscataway, NJ) and cleaned by Spin column chromatography using MicroSpin G-50 Columns (Amersham Biosciences). FINRISK WGA DNA samples were genotyped as genomic DNA samples in the study III and 59 WGA samples were included in the study of reproducibility and accuracy of the genotyping (IV) (Figure 12).

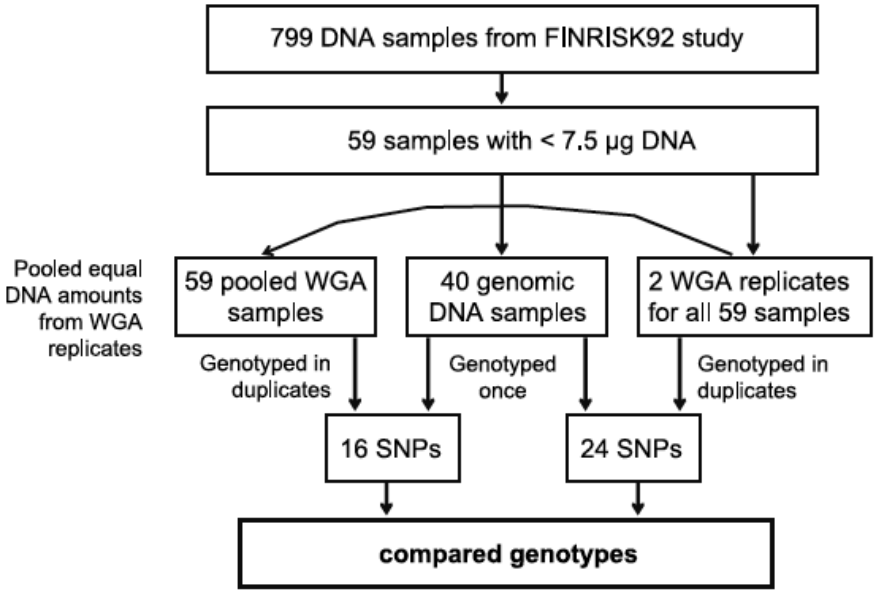

Figure 12. Study design of the whole genome amplification of DNA samples with low DNA yield. From the FINRISK study cohort, 59 low-yield DNA samples were amplified with whole genome amplification (WGA) in two replicates. Forty of the samples were also available for direct genotyping of the genomic DNA. These samples were genotyped with 24 SNPs and the genotypes were compared. In addition, 16 of the SNPs were genotyped using pooled samples of the WGA replicates.

4.4 Statistical methods

Haplotypes of the *USF1* region were estimated with Haploview 3.2 using default settings (Barrett et al. 2005) and the haplotype carrier status of the study subjects was determined from the haplotype-tagging properties of the SNP alleles (I-III).

Pearson's $\chi 2$-statistic was used to compare the allele frequencies between cases and non-cases (I, III).

Time-to-event analyses were performed with the Cox proportional hazards regression model (I, III). In the FINRISK analyses (III), a weighted Cox proportional hazards model, modified to account for the case-cohort sampling design, was used for risk analyses with the variance correction based on the published literature using SAS PHREG procedure (Prentice 1986; Lin and Wei

1989; Barlow 1994). Age represented the time measure in the models, which were adjusted for sex, hypertension, smoking, diabetes, total cholesterol to HDL ratio, and BMI. Individuals with prevalent CVD at the baseline were excluded from analyses of incident CVD (I, III). FINRISK analyses (III) were stratified for eastern and western Finland, and the FINRISK cohort was used as a covariate. In the analyses, the SNP genotypes were analyzed as two groups; the carriers of the minor alleles vs. the non-carriers of the minor alleles (I, III). In the FINRISK analyses, a grouping of carriers of the major alleles vs. non-carriers of the major alleles was also used.

Associations between *USF1* alleles and several quantitative variables were tested with general linear models (PROC GLM of SAS for Windows V8/9.1) (I, III) or with mixed longitudinal models (PROC MIXED of SAS for Windows V8) (I). The ULSAM dataset (I) was analyzed in groups comprising the whole population cohort, CVD patients, MetS patients, and healthy subjects free of CVD and MetS. In FINRISK (III), men and women were analyzed independently in the groups of incident CVD cases and in subcohort members free of CVD. Outliers (+/- 4SD) were removed before analysis, and HDL (I, III), triglyceride (I, III), LDL (I), TGs (I), BMI (I), waist grid (I), and systolic and diastolic blood pressures (I) were log-transformed to correct them to be normally distributed. Lp(a), fasting glucose, HOMA β-cell, and the clamp variables were corrected with rank procedure (I). Study subjects taking medication for their lipid levels, hypertension, or diabetes, were excluded from the analyses for these variables. To evaluate the significance of our findings, we permuted the genotype while retaining the phenotype data within sex/cohort groups, and repeated the same analyses that were performed with the actual dataset, recording the minimum p-value observed. We repeated this procedure 1,000 times and took the fifth percentile of these minimum p-values as the new multiple-testing corrected threshold for the p-values obtained with the original data.

Logistic regression models (PROC LOGISTIC of SAS for Windows V8/9.1) were used to test association of the *USF1* with atherosclerotic variables or sudden cardiac death in the HSDS study (II). Genotype-based and haplotype-based analyses were carried out on a combined data set of the HSDS 1981-2 series and 1991-2 series, and also on both separately, to establish how the associations observed in the combined data replicate in the two series. The genotypes were analyzed under a codominant model and in the haplotype analysis haplotype carriers were compared against non-carriers. Due to the skewed distribution of the quantitative measurements of atherosclerotic lesions, we modeled the risk of atherosclerosis with the multivariable ordinal regression model (cumulative logit model) where quartiles of the atherosclerotic variable were the ordinal outcome. In these analyses, BMI, age, and HSDS series were covariates in the models. The risk of sudden cardiac death was analyzed as a binary outcome in a standard logistic regression model, where age and

HSDS series were covariates. Statistical significance in the analyses was set at 0.05. The p-values were adjusted for multiple comparisons with a false discovery rate (FDR) implemented in the SAS MULTTEST procedure. Each adjustment procedure included one SNP or a haplotype, and all tested phenotypes.

5 RESULTS AND DISCUSSION

5.1 Genetic variants of *USF1* and risk factors for cardiovascular disease (I, III)

At the initiation of our work, the association between *USF1* and FCHL had just been described in exceptional Finnish dyslipidemic families (Pajukanta et al. 2004). Our aim was to examine whether variation at the *USF1* locus would be of more general importance for lipid phenotypes and other risk factors for cardiovascular disease at the population level. This was achieved through examination of *USF1* variants in two large prospective population cohorts.

The follow-up period of the ULSAM (I) population cohort included four medical examinations, where measurements for serum lipids, blood pressure, and body anthropometrics were made (Table 12).

Table 12. Characteristics of the ULSAM (I) cohort.

	Age 50 (n=2,322)	**Age 60 (n=1,860)**	**Age 70 (n=1,221)**	**Age 77 (n=839)**
	Mean (SD)			
Total cholesterol (mmol/l)[a]	6.87 (1.31)	6.31 (1.10)	5.82 (1.00)	5.48 (0.96)
LDL cholesterol (mmol/l)[a]	5.26 (1.24)	4.28 (0.96)	3.90 (0.89)	3.56 (0.84)
Triglycerides (mmol/l)[a]	1.93 (1.22)	1.85 (1.01)	1.45 (0.78)	1.36 (0.68)
HDL cholesterol (mmol/l)[a]	1.36 (0.40)	1.19 (0.34)	1.28 (0.35)	1.32 (0.33)
Blood pressure (mmHg)[b]	132/83 (17/11)	140/86 (19/9)	144/82 (17/9)	149/80 (21/10)
BMI (kg/m^2)	25.03 (3.22)	25.49 (3.28)	26.30 (3.43)	-
Waist circumference (cm)	87.86 (8.84)	90.99 (10.03)	94.82 (9.65)	95.60 (9.86)
ApoB-100 (g/l)[a]	1.24 (0.28)	-	1.03 (0.23)	-
ApoA-I (g/l)[a]	1.43 (0.25)	-	1.28 (0.23)	-
Lp(a) (U/l)[a]	245.9 (305.2)	-	305.3 (360.4)	-
Fasting glucose (mmol/l)	4.96 (0.61)	4.81 (0.58)	5.36 (0.58)	5.48 (0.58)

[a] Individuals with lipid lowering medication removed (n=166)

[b] Individuals using antihypertensive medication removed (n=466)

Variation at the *USF1* locus associated with longitudinally measured BMI, total cholesterol, HDL and LDL, either among participants with metabolic syndrome or with CVD (Table 13). Associations among individuals free of these conditions were, however, absent (Table 13).

Table 13. Association (p-values) of the *USF1* SNPs with longitudinally measured cardiovascular risk factors

Variable	***rs2073658***			***rs2774279***			***rs2516839***		
	MetS[a]	CVD[b]	Healthy[c]	MetS[a]	CVD[b]	Healthy[c]	MetS[a]	CVD[b]	Healthy[c]
BMI					0.05(+)			0.05(+)	
Waist									
TC	0.01(-)[d]								
LDL	0.02 (-)								
HDL				0.02(+)					

A - indicates that the minor allele associates with an unfavorable profile of the variable, a + indicates association of the minor allele with a favorable profile.

[a] Metabolic syndrome by the ATP III criteria

[b] CVD: Individuals with cardiovascular events

[c] Individuals with no CVD events or metabolic syndrome

[d] Significant finding after multiple testing correction as 5th percentile of 1,000 permutations

In addition, associations between SNP *rs2073658* and longitudinally measured lipoproteins related to FCHL were observed; minor allele of the SNP associated with higher ApoB-100 values among participants with MetS (p=0.01), and with higher Lp(a) values among CVD cases (p=0.02) and healthy individuals (p=0.03). Only the ApoB-100 association, however, remained significant after correction for multiple testing. Minor allele of SNP *rs2774279* associated with lower fasting glucose levels among participants with CVD (p=0.01, significant after correction for multiple testing).

A cross-sectional analysis of the data at 70 years, which compared SNP associations among four different clinical definitions of MetS, further supported a protective role of the minor allele (A) of *rs2774279*; The minor allele carriers had higher mean HDL (p=0.02 – p=0.06) and lower mean Lp(a) (p=0.005 – p=0.07) values compared to the non-carriers of the allele, independently of the MetS definition used.

Furthermore, the minor allele frequency of *rs2774279* was lower among those with MetS diagnosed at one or more surveys when compared to healthy individuals (0.29 vs. 0.33, p<0.01, significant after adjusting the significance level for multiple tests).

In the FINRISK (III) studies, lipid and body composition measurements of the follow-up cohorts were conducted at the baseline of the study (Table 14).

Table 14. Baseline characteristics of FINRISK (III) cardiovascular cases and the subcohort

	FR92+FR97 Males		FR92+FR97 Females	
Characteristic	**CVD cases[a]**	**Subcohort[b]**	**CVD cases[a]**	**Subcohort[b]**
	Mean (SD)			
Number of subjects	374	451	154	175
Total cholesterol (mmol/l)[c]	6.02 (1.09)	5.76 (0.99)	6.06 (1.15)	5.98 (1.03)
HDL cholesterol (mmol/l)[c]	1.17 (0.30)	1.27 (0.34)	1.37 (0.38)	1.56 (0.36)
Non-HDL cholesterol[c]	4.85 (1.09)	4.49 (1.00)	4.68 (1.19)	4.41 (1.03)
Triglycerides[c]	2.00 (1.14)	1.74 (1.02)	1.82 (1.05)	1.41 (0.81)
Waist to hip ratio	0.97 (0.06)	0.95 (0.06)	0.84 (0.07)	0.81 (0.06)
Body mass index (kg/m^2)	28.6 (3.6)	27.5 (3.9)	29.3 (4.9)	27.3 (4.6)

[a] Incident CVD cases (includes incident CHD cases and incident ischemic stroke cases)
[b] Subcohort members without CVD at baseline or during the follow up
[c] Persons with lipid lowering medication at baseline (n=44) were excluded from calculations

Although some suggestive associations between *USF1* and the measurements were observed in the FINRISK (III) subcohort (Table 15), only one remained significant after permutation correction for multiple testing; the G-allele of *rs2774276* was associated with a higher waist-to-hip ratio among men. Interestingly several associations were observed among participants with CVD (Table 15). Four *USF1* SNPs associated with total cholesterol values, and three SNPs with non-HDL cholesterol values, especially among men. SNP *rs2516839* associated with non-HDL cholesterol and total cholesterol values both among men and women, and also with BMI among women (Table 15).

Table 15. Association (p-values) between *USF1* SNP alleles and lipids, BMI, and waist-to-hip ratio in FINRISK (III) CVD case men and women

Marker	Allelea	Group	TC (mmol/l)	HDL (mmol/l)	Non-HDL cholesterol	TG (mmol/l)	BMI (kg/m2)	Waist-To-Hip ratio
rs10908821	G	M						S
		F						
rs2073658	C	M	0.02(-)		0.03(-)			
		F				0.03(-)		
	T	M	0.03(+)		0.02(+)	0.04(+)		
		F		S				S
rs2774276	C	M	0.01(+)		0.004(+)[b]			
		F						
	G	M						S
		F						
rs2516839	C	M	0.001(-)[b]		0.001(-)[b]			S
		F	0.02(-)		0.04(-)		0.006(-)[b]	
	T	M	0.04(+)		0.04(+)			
		F						
rs1556259	G	M	0.03(-)					
		F						
rs2774279	A	M						
		F						
	G	M	S		S			
		F		S				

A - indicates that allele associates with an unfavorable profile of the variable, a + indicates association with a favorable profile. M=males, F=females, S=suggestive association in the subcohort

[a] Persons carrying the allele are pooled together and analyzed against non-carriers

[b] p-value less than the critical value corrected for multiple testing

Several studies in dyslipidemic families have demonstrated linkage or association between *USF1* gene and FCHL (Pajukanta et al. 2004; Coon et al. 2005; Huertas-Vazquez et al. 2005; Lee et al. 2007), a complex dyslipidemia often implicated in patients with premature coronary heart disease (Genest et al. 1992). Our data suggests that variation at the *USF1* locus may also contribute to lipid phenotypes and other cardiovascular risk factors more widely at the population level.

The FCHL phenotype comprises various lipid component traits and several studies have assessed them in the presence of *USF1* variation (Figure 13, Table 16, and Table 17). Many of those studies have included polymorphisms *rs2073658* and *rs3737787*, two highly correlated SNPs with the strongest associations to FCHL and triglycerides in the first FCHL study by Pajukanta and colleagues (Pajukanta et al. 2004). In addition to FCHL, the common allele of the two SNPs has associated with higher levels of apolipoproteins and free fatty acids (Table 16). Furthermore, carriers of the common alleles of these risk SNPs have shown higher levels of serum

total cholesterol, triglycerides and LDL (Table 17). While many studies have reported significant associations with these lipid phenotypes, replication has at times been LDL (Table 17). While (Gibson et al. 2005; Hoffstedt et al. 2005; Ng et al. 2005; Zeggini et al. 2006; Kantartzis et al. 2007).

Most of the *USF1* studies have ascertained their study samples with specific criteria, such as presence of FCHL, CVD, diabetes, or obesity (Table 16 and Table 17). Our samples were Finnish (FINRISK, III) and Swedish (ULSAM, I) prospective population cohorts. In the ULSAM (I) cohort, longitudinal measurements of serum lipids and other cardiovascular risk factors were examined in relation to *USF1* variation. Contradictory to studies where the common allele (C) of *rs2073658* associated with a more unfavorable risk profile, in the ULSAM (I) the minor allele (T) of *rs2073658* associated with a higher total and LDL cholesterol, apolipoprotein B-100, and lipoprotein-a concentrations. The risk trend prevailed throughout the follow-up time of three decades. Interestingly, the ULSAM (I) associations were almost solely present among participants with MetS or CVD, suggesting that a certain pathophysiological state is required for the *USF1* gene effect to emerge. This was further indicated in the FINRISK (III) study sample where the common allele of *rs2073658*, and other SNPs, contributed to higher lipids in CVD cases, but not in the subcohort (Table 17).

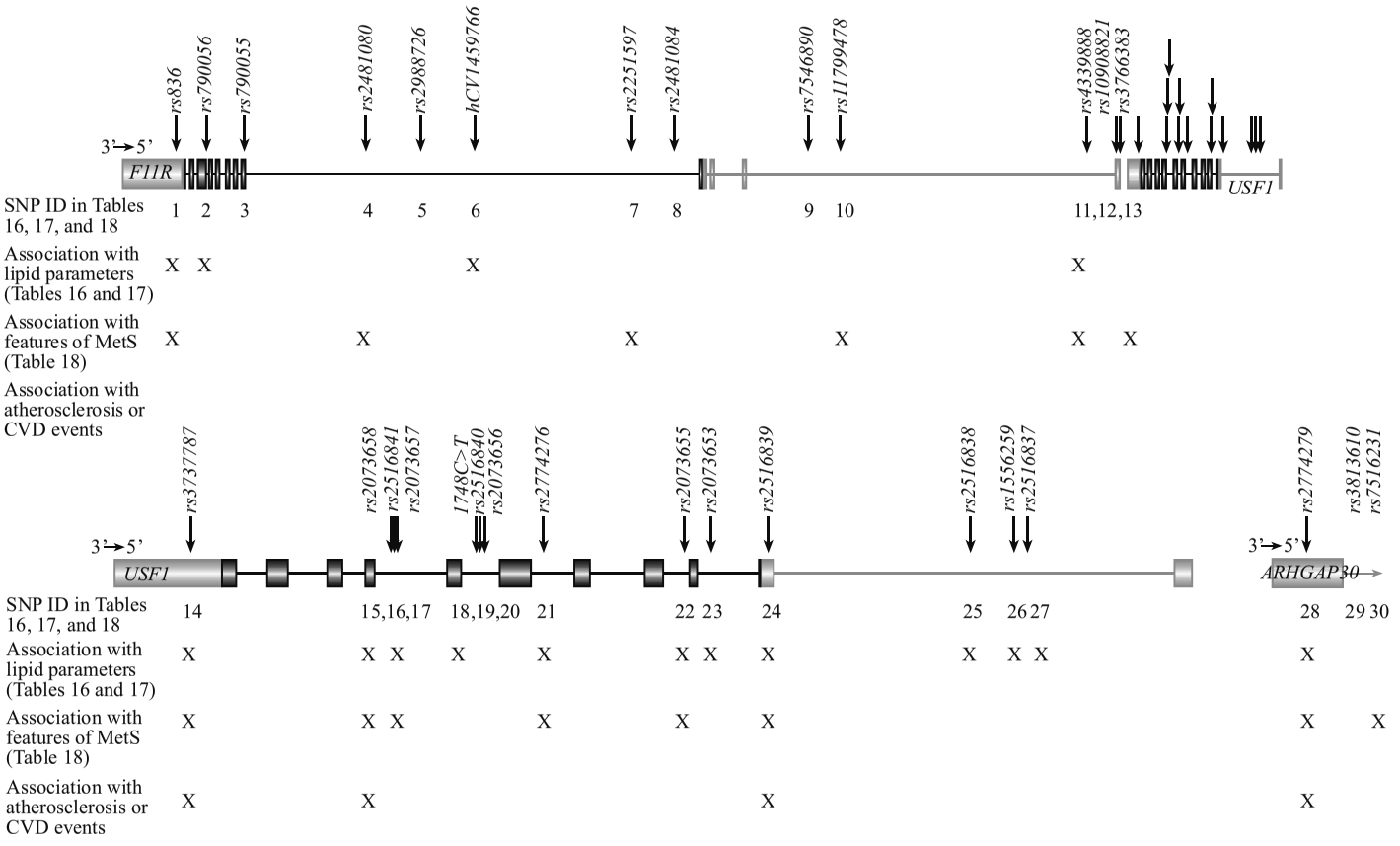

Figure 13. *USF1* SNPs indicated in Tables 16, 17, and 18. The transcribed regions of the genes are 43.8 kb (*F11R*), 6.7 kb (*USF1*), and 23.0 kb (*ARHGAP30*). Distance from the *USF1* transcript is 0.26 kb to *F11R* and 0.97 kb to *ARHGAP30*.

Table 16. Association of *USF1* SNPs with FCHL component traits.

Study sample (and SNPs studied[a])	Sex	Ascertainment or health status	apoB/ apoA1/ apoE	lp(a)	Free Fatty acids	REF
ULSAM, longitudinal population cohort of Swedish men investigated at the ages of 50, 60, 70, and 77. (15, 24, 28)	M	MetS and/or CVD	15-	15-28+	NA	(Auro et al. 2007)
ULSAM, longitudinal population cohort of Swedish men investigated at the ages of 50, 60, 70, and 77. (15, 24, 28)	M	Healthy	-	15-	NA	
EARSII, a multicenter study of healthy young (18-28) men, whose father had a premature MI, and their controls. (18, 22, 23)	M	Family history of premature MI. BMI*SNP interaction.	22?	NA	NA	(Putt et al. 2004)
Dutch FCHL families. (14, 15)	M+F	FCHL families	14+? 15+?	NA	NA	(van der Vleuten et al. 2007)
	M		-	NA	NA	
Finnish FHCL families. (1-3, 6, 11-17, 19, 23-26)	M	FCHL families (proband with CHD)	14+ 15+	NA	NA	(Pajukanta et al. 2004)
	M+F		14+ 15+	NA	NA	
Dutch FCHL Families. (14)	M	FCHL families	14+	NA	NA	(Lee et al. 2007)
	F		-	NA	NA	
Quebec Family Study cohort, representing a mixture of random sampling and ascertainment through obese probands. (14, 15, 24)	M	Random sampling or obesity	-	NA	NA	(Choquette et al. 2007)
	F		-	NA	NA	
Healthy, obese, Swedish women. (14, 15)	F	Obesity	-	NA	-	(Hoffstedt et al. 2005)
Nondiabetic, healthy Caucasians. (14, 15)	M+F	Healthy	NA	NA	14+ 15+	(Kantartzis et al. 2007)
Type 2 diabetes case-control study, French Caucasians. (4, 14-16, 21, 23, 25, 26)	M+F, M, F	Normoglycemic	-	NA	NA	(Gibson et al. 2005)
Utah families, nondiabetic family members. (14, 16, 19, 24-26)	M+F?	Evidence for 1q linkage, nondiabetic family members	NA	NA	-	(Zeggini et al. 2006)
Unrelated Arkansas caucasians with normal glucose tolerance test. (14, 16, 19, 24-26)	M+F?	Age 18-50	NA	NA	-	

[a] The rs-numbers of the studied SNPs are given in Figure 13. A + indicates that the minor allele associates with a more favorable risk profile. A - indicates that the minor allele associates with an unfavorable risk profile. If the minor allele differs between populations, the one defined in the Finnish population is used. NA=not available, M=males, F=females.

Table 17. *USF1* SNPs associated with serum lipid levels.

Study sample (and SNPs studied[a])	Sex	Ascertainment or health status	TC	TG	HDL	LDL[b]	REF
ULSAM, longitudinal population cohort of Swedish men investigated at the ages of 50, 60, 70, and 77. (15, 24, 28)	M	MetS and/or CVD	15-	NA	28+	15-	(Auro et al. 2007)
FINRISK, Finnish population follow-up cohort of 25-74 year old participants. (12, 15, 21, 24, 26, 28)	M	CVD	15+ 21- 24- 26-	15+	-	NA	(Komulainen et al. 2006; Kristiansson 2007)
	F		24-	15+	-	NA	
FINRISK, Finnish population follow-up cohort of 25-74 year old participants. (12, 15, 21, 24, 26, 28)	M	healthy	28-	-	-	NA	
	F		-	-	15+ 28-	NA	
24 extended Mexican FCHL families with a history of premature CHD. (1, 6, 8, 11, 13, 15, 20, 22, 25, 28, 29)	M+F	FCHL families (proband with CHD)	NA	6+ 14+ 15+	NA	NA	(Huertas-Vazquez et al. 2005)
Dutch FCHL families. (14, 15)	M+F	FCHL families	14+? 15+?	-	NA	-	(van der Vleuten et al. 2007)
Finnish FHCL families. (1-3, 6, 11-17, 19, 23-26)	M	FCHL families (proband with CHD)	14+ 15+	1 2 6 11 14+ 15+ 24 25	NA	14+ 15+	(Pajukanta et al. 2004)
	M+F		14+ 15+	14+ 15+	NA	14+ 15+	
Dutch FCHL Families. (14)	M	FCHL families	14+	14+	NA	NA	(Lee et al. 2007)
EARSII, a study of healthy young (18-28) men, whose father had a premature MI, and their controls. (18, 22, 23)	M	Family history of MI. BMI*SNP interaction.	-	-	-	18? 23?	(Putt et al. 2004)
Utah pedigrees ascertained for early death due to CHD, early strokes, or early onset hypertension. (14, 15)	M+F	CHD, strokes, hypertension	NA	14+	NA	14+ 15+	(Coon et al. 2005)
GeneBank Cohort, sequential subjects age 18 years or older undergoing diagnostic cardiac catheterization. (14)	M	CAD cases	-	14+	NA	NA	(Lee et al. 2007)
	F		-	14-	NA	NA	
CARDIA, prospective US cohort of young adults. (14, 21, 25-27)	M+F	Mixed	14+	-	-	14+ 27-	(Reiner et al. 2007)
CHS, a prospective US population-based cohort of older men and women. (22, 25-27)	M+F	Mixed	22+ in obese	25+	-	-	
Quebec Family Study cohort, a mixture of random sampling and ascertainment through obese probands. (14, 15, 24)	M	Random sampling or obesity	-	-	-	24-	(Choquette et al. 2007)
Utah families, nondiabetic family members. (14 ,16, 19, 24-26)	M+F?	Evidence for 1q linkage, nondiabetic family members	16+ 24+ 25-	14- 24+	NA	NA	(Zeggini et al. 2006)

[a] The rs-numbers of the studied SNPs are given in Figure 13. A + indicates that the minor allele associates with a more favorable risk profile. A - indicates that the minor allele associates with an unfavorable risk profile. If the minor allele differs between populations, the one defined in the Finnish population is used. NA=not available, M=males, F=females. [b] LDL cholesterol level or LDL peak particle size.

In addition to the involvement of *USF1* in lipid phenotypes, our population cohorts also revealed the association of *USF1* SNPs with some other important CVD risk factors (Table 18). Both in FINRISK (III) and in ULSAM (I), the SNP *rs2516839* associated with the BMI or waist-to-hip ratio (Table 18). The trend of the effect of the rare allele of *rs2516839* in ULSAM (III) was, however, discordant to other studies (Table 18). *Rs2774279* also associated with BMI in the ULSAM (I) study sample of Swedish men. This association was also present in subsequent analyses of the FINRISK (III) cohort where SNP associations were analyzed under a codominant model instead of the dominant and recessive models (Kristiansson 2007) (Table 18). Furthermore, the association of *rs2774279* with BMI displayed a sex-specific effect, as the rare allele correlated with a favorable risk profile among men and an unfavorable one among women (Table 18) (Kristiansson 2007). While the possibility of a spurious finding exists, these types of discordances, which have also been reported in other studies (Lee et al. 2007; Reiner et al. 2007), may reflect the complicated, interacting nature of the *USF1* effect. Suggestive evidence of gene-sex, gene-gene, gene-age, and other interactions exists for *USF1* (Putt et al. 2004; Coon et al. 2005; Choquette et al. 2007; Kantartzis et al. 2007; Lee et al. 2007; Reiner et al. 2007). Reports of associations specific to groups of study subjects with certain pathophysiological state, such as obesity, MetS, or CVD, further add to the complexity.

We observed *USF1* to associate with MetS in the ULSAM (I) cohort, an association also suggested in other studies (Table 18). Although we, in concordance to some other studies (Table 18), also detected *USF1* to associate with fasting glucose values, the contribution of *USF1* to type 2 diabetes remains unclear. Initial associations (Ng et al. 2005; Zeggini et al. 2006) have failed to be replicated in other populations or subsequent studies (Gibson et al. 2005; Ng et al. 2005; Zeggini et al. 2006; Reiner et al. 2007).

Table 18. *USF1* SNPs associated with body composition measures, diabetes, and MetS.

Study sample (and SNPs studied[a])	Sex	Ascertainment or health status	Body composition	Glucose/insulin/HOMA	Diabetes	MetS	REF
ULSAM, longitudinal population cohort of Swedish men investigated at the ages of 50, 60, 70, and 77. (15, 24, 28)	M	MetS and/or CVD	24+ 28+	28+	NA	15- 28+	(Auro et al. 2007)
FINRISK, Finnish population follow-up cohort of 25-74 year old participants. (12, 15, 21, 24, 26, 28)	M	CVD	-	NA	NA	NA	(Komulainen et al. 2006; Kristiansson 2007)
	F		15+ 24-	NA	NA	NA	
FINRISK, Finnish population follow-up cohort of 25-74 year old participants. (12, 15, 21, 24, 26, 28)	M	Healthy	12- 21- 24- 28+	NA	NA	NA	
	F		15+ 28-	NA	NA	NA	
Quebec Family Study cohort, representing a mixture of random sampling and ascertainment through obese probands. (14, 15, 24)	M	Random sampling or obesity	14+ 15+ 24-	-	NA	NA	(Choquette et al. 2007)
	F		14+ 15+	-	NA	NA	
EARSII, a multicenter study of healthy young (18-28) men, whose fathers had a premature MI, and their controls. (18, 22, 23)	M	Family history of premature MI. BMI*SNP interaction.	-	22?	NA	NA	(Putt et al. 2004)
Dutch FCHL Families (14)	M	FCHL families	-	NA	NA	14+	(Lee et al. 2007)
GeneBank Cohort, sequential subjects age 18 years or older undergoing diagnostic cardiac catheterization (14)	M	CAD cases	14+	-	NA	NA	
	F		14-	-	NA	NA	
Nondiabetic, healthy Caucasians (14, 15)	M+F	Healthy	-	14+ 15+	NA	NA	(Kantartzis et al. 2007)
Multiethnic case-control study, type 2 diabetes (1, 3-5, 7-10, 11, 13, 14, 19, 20, 23-26, 28-30)	M+F ?	Evidence for 1q linkage, diabetes cases and controls	NA	NA	1C+ 4C- 7C- 10C- 11C- 13G- 24- 30A-	NA	(Zeggini et al. 2006)
Hong Kong Family Diabetes Study, families and type 2 diabetic and/or metabolic syndrome hospital cases and controls (14, 16, 24)	M+F ?	Diabetes	NA	-	14+ 16-	14+	(Ng et al. 2005)

[a] The rs-numbers of the studied SNPs are given in Figure 13. A + indicates that the minor allele associates with a more favorable risk profile. A - indicates that the minor allele associates with an unfavorable risk profile. If the minor allele differs between populations, the one defined in the Finnish population is used.
NA=not available, M=males, F=females.

5.2 *USF1* and atherosclerosis (II)

Cholesterol, triglycerides, and other atherogenic components that circulate in the blood or are present in arterial vessel walls, crucially contribute to atherosclerosis and its complications, such as myocardial infarction and stroke. *USF1* associated with risk factors for cardiovascular disease in our studies, as well as in other studies. We thus sought to relate variation at the gene locus to the extent and severity of atherosclerosis at the level of the vessel walls of the coronary arteries and abdominal aorta. To accomplish this, we studied the vessel walls of 700 middle-aged men from the HSDS (II) study where medicolegal autopsies provided detailed data on several types of atherosclerotic changes on the arteries of the study subjects (Table 19). Two HSDS study sample series collected in 1981-82 (n=400) and 1991-92 (n=300) were analyzed.

Table 19. Characteristics of the Helsinki Sudden Death Study (HSDS) (II) sample

	1981-82 series (n=400)	**1991-92 series (n=300)**
Age , years±SD	53.8±9.5	52.1±9.6
Body mass index, kg/m^2±SD	24.2±4.6	25.1±5.0
Cause of death, n (%)		
Sudden cardiac death	150 (37.5)	80 (27)
Other disease	98 (24.5)	100 (33)
Non-natural deaths (accidents, suicides)	152 (38)	120 (40)
Presence of coronary atherosclerotic changes, n (%)		
Fatty streaks	381 (95)	282 (94)
Fibrous lesions	346 (87)	232 (77)
Complicated lesions	195 (49)	101 (34)
Calcified lesions	317 (79)	205 (68)
> 50% coronary narrowing in more than one coronary artery	120 (30)	38 (17)

Genotype and haplotype analyses of *USF1* in the HSDS (II) study samples revealed a single SNP, which consistently associated with several kinds of atherosclerotic lesions in the arteries of the HSDS study subjects. Study subjects with two copies of the *rs2516839* T-allele had mainly larger areas of atherosclerotic changes than those with only one copy of the T-allele. Conversely, the CC genotype associated with a protective effect (Table 20). The risk effect of the T-allele on advanced

atherosclerosis and calcification was present both in the coronary arteries and in the abdominal aorta of the study subjects (Table 20).

Table 20. Association between *USF1 rs2516839* genotype and atherosclerosis in coronary arteries and abdominal aorta

	Percentage area of atherosclerosis lesion type, mean±SD			TT vs. CC	
	CC (n=100)	CT (n=304)	TT (n=253)	OR (95% CI)[a]	p-value[b]
Coronary arteries					
Fatty streaks	8.82±7.04	8.41±7.40	8.17±6.95	0.79 (0.52-1.21)	0.28 (0.29)
Fibrous lesions	5.73±5.34	7.67±8.91	7.89±7.22	1.65 (1.08-2.52)	0.02 (0.05)
Complicated lesions	1.33±2.81	2.86±7.30	2.99±5.43	1.82 (1.11-2.98)	0.02 (0.05)
Total lesion area	14.40±8.78	17.04±13.12	16.81±11.22	1.30 (0.85-1.99)	0.23 (0.26)
Advanced atherosclerosis[c]	6.78±6.29	9.81±11.64	9.94±8.86	1.72 (1.12-2.64)	0.01 (0.04)
Proportion of advanced lesion area out of the total lesion area	0.50±0.32	0.57±0.31	0.60±0.30	1.91 (1.24-2.94)	0.003 (0.02)
Calcified lesions	2.36±3.69	4.03±6.41	5.40±7.24	2.80 (1.80-4.35)	<0.001 (<0.001)
Abdominal aorta					
Fatty streaks	11.15±9.12	14.27±10.91	12.95±10.26	1.50 (0.98-2.28)	0.06 (0.09)
Fibrous lesions	5.56±5.33	6.48±6.19	7.12±6.51	1.44 (0.94-2.20)	0.09 (0.11)
Complicated lesions	5.70±10.71	6.37±9.27	7.02±10.52	1.49 (0.95-2.34)	0.08 (0.10)
Total lesion area	22.40±16.60	26.97±16.96	26.88±17.29	1.59 (1.04-2.43)	0.03 (0.06)
Advanced atherosclerosis[c]	11.25±13.16	12.86±12.80	14.17±13.57	1.65 (1.07-2.56)	0.02 (0.05)
Proportion of advanced lesion area out of the total lesion area	0.42±0.26	0.43±0.27	0.47±0.27	1.26 (0.82-1.94)	0.29 (0.29)
Calcified lesions	3.29±5.20	3.85±6.12	4.34±6.13	1.60 (1.03-2.48)	0.04 (0.06)

SD: Standard deviation; OR: Odds ratio; CI: Confidence interval

[a] Odds ratio and 95% confidence interval from ordinal logistic regression analysis. The variables were analyzed as quartiles with an ordinal regression model due to their skewed distribution. See publication II.

[b] p-value adjusted for multiple comparisons in parenthesis

[c] Plaque area covered either by fibrotic or complicated lesion

The effect of *rs2516839* was observed independently in the two HSDS (II) autopsy series: Carriers of two risk alleles were 2.4 times as likely to have more severely calcified coronary arteries in the 1981-82 series (OR 2.39, 95% CI 1.30-4.40) and 3.2 times as likely in the 1991-92 series (OR 3.20, 95% CI 1.68-6.09), than non-carriers of the allele (Figure 14).

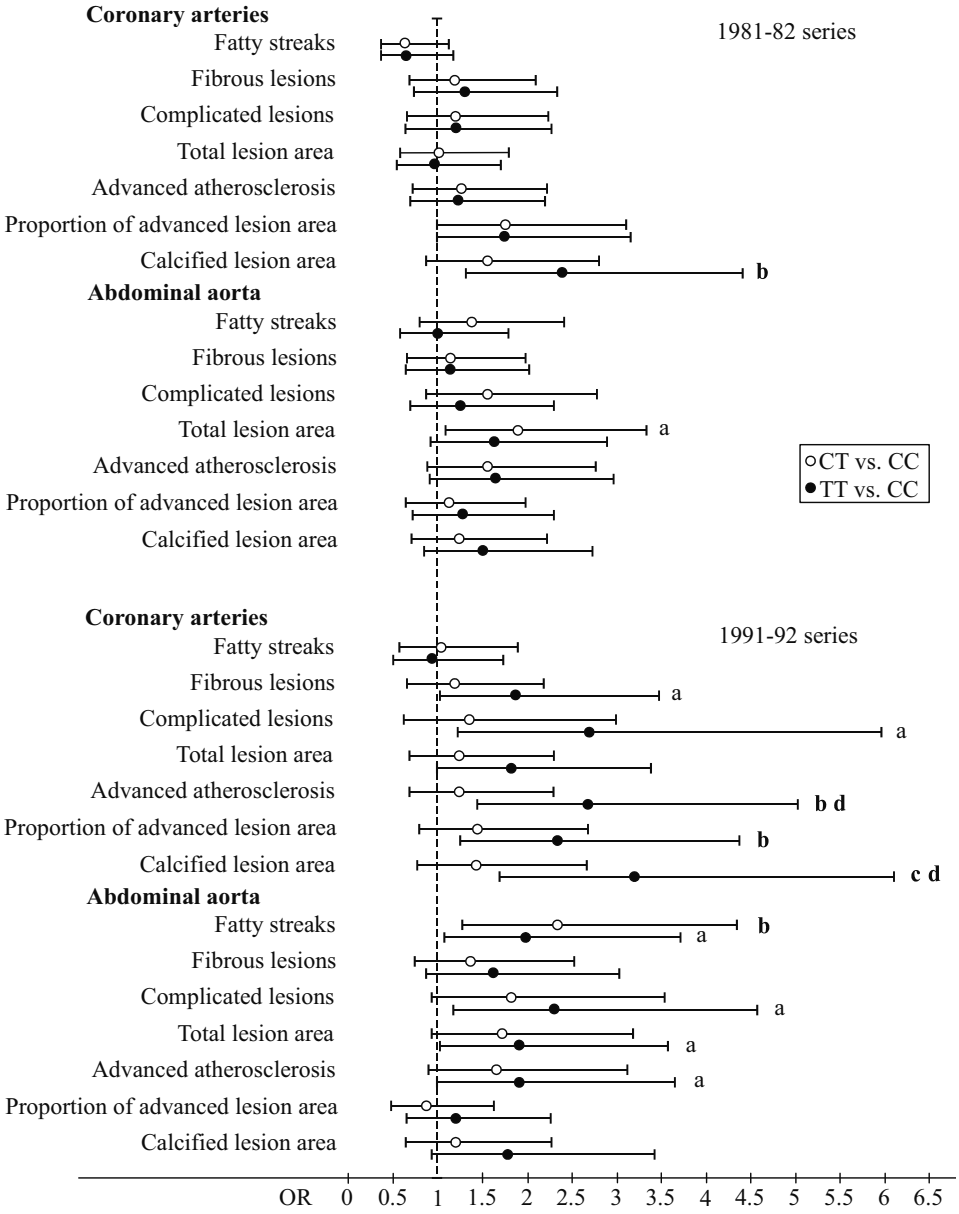

Figure 14. Association of *rs2516839* with calcification of the arteries in the two HSDS (II) autopsy series. The figure shows odds ratios and their confidence intervals from the ordinal logistic regression analysis which compared the CT and TT genotypes to the CC genotype. Advanced atherosclerosis = plaque area covered either by fibrotic or complicated lesion. [a] $p<0.05$, [b] $p<0.01$, [c] $p<0.001$, [d] Multiple testing adjusted $p<0.05$.

In addition to *rs2516839*, SNPs *rs2073658* and *rs2774279* showed statistically significant association with the amount of calcification in the coronary arteries. These associations were, however, non-significant in regression models adjusted for SNP *rs2516839*. The associations obtained from the analyses of *rs2516839* were also stronger than the few suggestive associations seen in haplotype analyses.

We and others have presented data on the effect of *USF1* on several cardiovascular risk factors, most prominently on lipid phenotypes (Table 16 and Table 17). These lipids participate in atherogenic processes leading to the development of plaques in the arteries and thus we hypothesized that genetic variation at the *USF1* locus could associate with severity of coronary atherosclerosis. We obtained evidence to support our hypothesis by studying coronary arteries and abdominal aortas of Finnish middle-aged men of the Helsinki Sudden Death Study (II).

Rs2516839 consistently associated with several kinds of atherosclerotic lesions in the arteries of the HSDS (II) study subjects (Table 20 and Figure 14) and the strongest association involved coronary calcification. A study by Reiner and colleagues also identified a link between *USF1* and coronary artery calcium (CAC): the rare allele of *rs3737787* was associated with a decreased risk of CAC in otherwise healthy young adults (Reiner et al. 2007). Their study included a SNP highly correlated with our *rs2516839*, but they reported no association between CAC and that particular SNP. In another study, evidence of *rs3737787* contributing to coronary stenosis in a prospective clinical cohort was absent (Lee et al. 2007).

Rs2516839 was undeniably the most significantly associated *USF1* polymorphism in our autopsy study sample, evidenced by the non-significance of the other SNPs in an analysis model where *rs2516839* was present. Although this could explain the lack of association in the study of Lee and colleagues where only *rs3737787* was investigated (Lee et al. 2007), it is contradictory to the study of Reiner and colleagues (Reiner et al. 2007). While several functional alleles at the *USF1* locus could exist, their identification is obscured by the high LD present in the region (Pajukanta et al. 2004; Reiner et al. 2007). Reiner and colleagues examined the nature of this LD and concluded that in their study the true susceptibility effect on mortality may be attributable, at least in part, to neighboring polymorphisms in other genes (Reiner et al. 2007). For atherosclerosis this hypothesis is particularly interesting: next to *USF1* resides a 42 kb gene encoding F11-receptor (F11R), which is expressed on the surface of platelets and localized within tight junctions of endothelial cells (EC). The gene was recently implicated in atherothrombosis in a study where the overexpression of F11R was present in atherosclerotic plaques (Babinska et al. 2002; Babinska et al. 2007).

Still, as *USF1* has a role in the regulation of several genes encoding components of lipid and inflammation pathways, its contribution to atherosclerosis remains plausible.

5.3 *USF1*, sudden cardiac death, and the incidence and prevalence of cardiovascular disease (I, II, III)

After finding evidence on the effect of the *USF1* locus on levels of CVD risk factors and atherosclerosis, we hypothesized that this effect might be observed as an occurrence of acute CVD events and death.

In the HSDS (II) study sample where *USF1* SNP *rs2516839* associated with atherosclerosis at the vessel wall level, the risk allele also associated with sudden cardiac death. The carriers of a TT genotype had a 2-fold risk for SCD compared to that of the carriers of the protective CC genotype (TT; OR 2.10, 95% CI 1.17-3.75, p=0.01, and CT; OR 1.92, 95% CI 1.09-3.39, p=0.02). Part, but not all, of the increased risk of SCD associated with the SNP was due to the SNP's contribution to the development of severe coronary artery disease characterized by coronary calcification.

Among the FINRISK (III) female, but not male, study subjects, carriers of the minor allele (A) of *rs2774279* were at a lower risk of belonging to the group of study subjects with a baseline history of a cardiovascular event (AA vs. GG: OR 0.51, 95% CI 0.26–0.99 and AG vs. GG: OR 0.71, 95% CI 0.51–1.00, p=0.05) (Kristiansson 2007). In addition, female carriers of the minor allele of *rs2073658* (T) had a higher risk of a baseline CVD history (TT vs. CC: OR 2.01, 95% CI 1.06–3.83 and CT vs. TT: OR 1.42, 95% CI 1.03–1.96, p=0.03) (Kristiansson 2007).

A relation between variation at the *USF1* locus and prevalent CVD was thus suggested by our results and we further extended our analyses of *USF1* to incidence of CVD with our prospectively followed population cohorts. Among the FINRISK (III) women, the two *USF1* SNPs that were associated with prevalent CVD also associated with the incidence of CVD during the follow-up period (Figure 15). Interestingly, these SNPs also associated with mortality; carriers of the *rs2073658* T-allele and carriers of the *rs2774279* G-allele had an increased risk of all-cause mortality compared to non-carriers of the allele (*rs2073658* T-allele: HR 2.52, 95% CI 1.46–4.35, p=0.0009; *rs2774279* G-allele: HR 4.43, 95% CI 1.58–12.40, p=0.005).

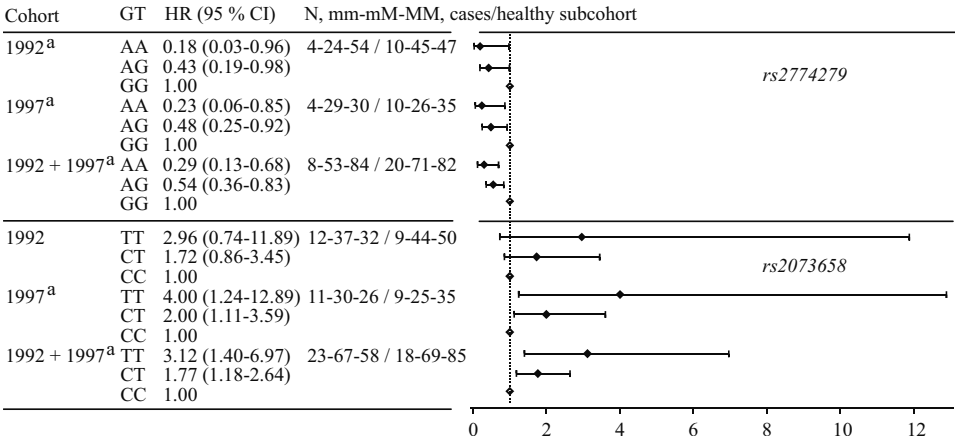

Figure 15. *USF1* and incidence of cardiovascular event among FINRISK (III) females.

[a] p<0.05, HR=hazard ratio, CI=confidence interval, mm=homozygote of the minor allele, mM=heterozygote, MM=homozygote of the major allele. A codominant model for the genotypic effect was used in the analysis.

Among FINRISK (III) men, association between *USF1* and CVD was absent. When a possibly more genetically predisposed group of FINRISK men was examined, however, a suggestive association between the SNP *rs2516839* and incident CVD was observed. Among those men, whose first degree relatives had a history of CVD, carriers of the minor allele of *rs2516839* were at a lower risk of CVD during the follow-up (CC vs. TT: HR 0.52, 95% CI 0.28–0.96, p=0.04 and CT vs. TT: HR 0.63, 95% CI 0.43–0.93, p=0.02) (Kristiansson 2007).

Analyses of the ULSAM (I) study sample men also failed to show any significant association between *USF1* variation and incidence of CVD.

Our data on the population cohorts supported the role of *USF1* alleles in lipid metabolism, atherosclerosis, and also contributing in a few other cardiovascular risk factors. In addition we obtained evidence for how the risk increasing effects of *USF1* alleles eventually play a part in sudden cardiac death or in incidence of CVD.

In the HSDS (II) study sample the atherosclerosis-associated risk allele also associated with sudden cardiac death among the men. The same allele associated with incidence of CVD among FINRISK (III) men, but only after a prospectively more genetically predisposed group of men was studied. *USF1* alleles failed to show any differential effect on the risk of CVD among ULSAM (I) men.

The FINRISK (III) study sample also included women, and among them the relation between *USF1* alleles and CVD was more evident than among men. Two *USF1* polymorphisms associated with the incidence of CVD. Carriers of the A-allele of *rs2774279* had a lower risk of CVD compared to the non-carriers of the allele. This protective effect of the allele was observed in both FINRISK cohorts (1992 and 1997) and also among the baseline CVD cases. In addition, the T allele of *rs2073658*, associated with an increased risk for CVD, raising an interesting question of a possible sex-specificity of the *USF1* effect, since the rare T allele has in many studies showed a protective effect especially among men (Table 17).

Since most of the SNPs genotyped in the three study samples (I, II, III) each tag one *USF1* haplotype, the SNP associations also represented haplotype associations. *USF1* SNPs *rs2073658*, *rs2516839*, and *rs2774279* were genotyped in all three study samples (I, II, and III), and their haplotype frequencies in those study samples were similar. The frequency of CCG was 38% (I), 39% (II), and 38% (III), frequency of CTA was 31% (I), 28% (II), and 29% (III), and frequency of TTG was 31% (I), 33% (II), and 34% (III). These frequencies differed slightly from the haplotype estimates in the SeattleSNPs database (UW-FHCRC European SNPs, 47 individuals) (SeattleSNPs, Variation Discovery Resource). SeattleSNP frequencies were 49% (CCG), 28% (CTA), and 17% (TTG). While SNPs *rs2073658* and *rs2774279*, and thus haplotypes TTG and CTA, associated with CVD among FINRISK (III) women, it was SNP *rs2516839*, which associated with sudden cardiac death (II) and CVD (III) among men. *Rs2516839* does not tag a single haplotype of *USF1*; instead, the risk-associated T-allele of the SNP is present in the haplotypes TTG and CTA. The haplotype CTA, which associated with a lower risk of CVD among women, thus presented a risk effect among men.

Only a few other studies have examined *USF1* on CVD (Lee et al. 2007; Reiner et al. 2007). Lee and colleagues examined *rs3737787*, a *USF1* SNP in high correlation with *rs2073658*, and in concordance with our results found a significant sex-specific effect for the SNP alleles in lipid and BMI analyses (Lee et al. 2007) (Table 17 and Table 18). They failed, however, to observe any SNP association with MI or revascularization. Lack of association was also present in a study by Reiner and colleagues in which *USF1* SNPs were analyzed in the incidence of CVD (Reiner et al. 2007). They, however, observed an association of *USF1 rs3737787* with risk of all-cause mortality. The risk allele was the same as the CVD risk increasing allele in our FINRISK (III) women.

5.4 Whole genome amplification of the low DNA yield samples (IV)

Some of the FINRISK (III) study samples we used had a low yield of DNA and were thus in danger of being excluded from the analyses. To avoid bias introduced by such exclusion, we amplified these DNA with whole genome amplification and evaluated whether the produced WGA genotypes were of good quality.

When 59 low-yield DNA samples were whole-genome amplified in replicate and then genotyped with 24 SNPs (16 using the MassARRAY System and 8 using the microarray system), multiple genotype discrepancies were observed for 13 of the samples (Table 21). These discrepancies were detected both between WGA replicates and between a WGA product and genomic DNA, and most were due to a drop-out of one of the alleles in heterozygote WGA samples. More heterogeneous allele ratios were observed for these 13 samples, and allele calling problems, low signal intensities, and genotypes escaping from genotype clusters were common (Figure 16). Presence of genotype discrepancies did not correlate with the DNA amount or purity, although higher starting DNA concentrations for WGA amplification showed a higher genotype call rates.

Table 21. Distribution of genotype discrepancies for the 13 problematic WGA samples

	Number of discrepant SNPs			**DNA concentration ng/µl**
Sample	MassARRAY	Microarray	Total	
1	2		2	7
2	2		2	11
3		2	2	90
4	1	2	3	31
5	1	2	3	>0
6	3	2	5	1
7	5		5	13
8	4	2	6	23
9	6		6	30
10	5	2	7	3
11	7	1	8	18
12	8	1	9	>0
13	5	6	11	2

Samples genotyped on a total of 24 SNPs. Genotype comparisons were made between WGA replicates of the same sample and genomic DNA.

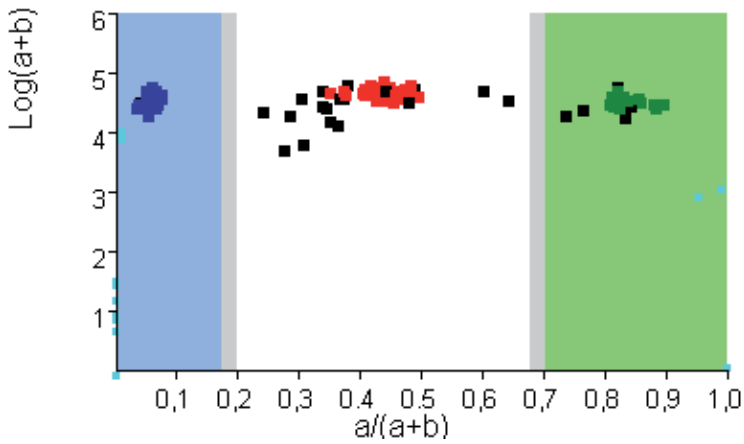

Figure 16. An example of genotype clusters for WGA samples. Clusters are for microarray genotyping of the SNP *rs2073658* for the 59 WGA DNA samples. The 13 samples that show multiple genotype discrepancies are shown as black dots, while the 3 genotype clusters for the other 46 samples are shown as colored dots. Two dots are present for each sample since genotyping was done in duplicates.

To evaluate whether pooling of the WGA replicates improves genotyping quality, equal amounts of the two WGA replicates for each sample were combined into one genotyping sample, which was then genotyped in duplicate with 16 SNPs using the MassARRAY System. The number of genotyping discrepancies was substantially reduced in the pooled samples (Table 22) and genotyping success rate was greater than 98%.

Table 22. Comparison of WGA-genomic genotype discrepancies obtained for single WGA samples and pooled WGA samples

	Number of samples with discrepancies	**Total number of discrepancies**	**Number of unique discrepancies**
WGA single reactions	12 / 40 (30%)	97 / 3,148 (3.1%)	60 / 847 (7.1%)
WGA pool	4 / 40 (10%)	12 / 1,125 (1.1%)	7 / 572 (1.2%)

Genotypes for genomic DNA were only available for 40 out of 59 samples.
Total number of discrepancies counted from 2 genotyping attempts for each replicate WGA-sample (i.e., a total of 4 genotyping attempts per sample for WGA single reactions, and 2 genotyping attempts for WGA pool), and one genotyping attempt for genomic DNA sample.
Number of unique discrepancies: each SNP-sample comparison counted once.

In general, MPRCA-WGA appeared to be a feasible approach for amplification of low-yield DNA samples prior to genotyping in studies where sample exclusion on the basis of their DNA yield could introduce a bias into the analyses. We studied genotypes from genomic DNA samples and from two replicates of WGA samples produced by two genotyping platforms and twenty-four SNPs. For a majority of the 59 samples studied, we obtained consistent WGA genotypes with high success rates. Thirteen samples, however, showed various genotype discrepancies. Furthermore, when we searched for characteristics common to these samples, such as correlation between the presence of discrepancies and DNA purity or concentration, and WGA DNA integrity or quantity, which could be used to distinguish them from other samples a priori, we failed to find any.

Although the problematic samples giving genotype discrepancies could not be detected before genotyping with standard laboratory quality control methods, our results suggested that enhanced quality control at the time of the genotyping could identify such samples; these samples were often identified as having more heterogeneous allele ratios, allele calling problems, low signal intensities, and genotypes escaping from genotype clusters.

Combining two WGA reactions into one pooled DNA sample for genotyping markedly reduced the number of discrepancies and samples showing them. This was in agreement with a previous study where the pooling of WGA products significantly reduced genotyping inconsistencies (Rook et al. 2004). Even after a pooling of WGA replicates, some inconsistency was observable in the genotypes produced, especially in the form of allele drop-outs. Thus in addition to pooling of WGA samples, increased quality control is required to complement for the higher rate of genotyping errors.

5.5 General discussion on *USF1* and the risk of CVD

We obtained evidence that variation at the *USF1* locus affects features of several CVD risk factors and contributes to atherosclerosis and to its symptoms; sudden cardiac death and prevalent and incident CVD. Our data enabled us to examine the effect of *USF1* at the population level and in a prospective manner (I, III), instead of in special families or study samples ascertained with specific criteria. In addition, we could investigate the association of *USF1* alleles with atherosclerosis using detailed data on different types of lesions at the vessel wall level (II).

Our data and methods had some limitations; while in the ULSAM (I) study sample lipid and anthropometrics measurements were taken repeatedly during the follow-up period, in FINRISK (III) only one baseline measurement was obtained. Thus, some

FINRISK CVD events occurred as much as ten years after the baseline measurements and changes in the risk factors during that time could not be accounted for in the analyses models. Furthermore, plasma lipid levels were unavailable in the HSDS (II) study samples and so we could not directly relate atherosclerotic lesion areas to lipid levels. The FINRISK (III) analysis of a particularly interesting lipid measurement for *USF1*, triglycerides, was biased due to measurements taken after only 4 hours of fasting. This possibly explains the lack of association in our study compared to others. Another limitation could arise from our definition of CVD; the diagnosis of CVD is dependent on several factors, including the point in time when the patient seeks medical care where CVD can be diagnosed. Our study samples may include persons with severe atherosclerosis, who are still considered "healthy" because their disease is absent from the national registers. Furthermore, by combining ischemic strokes with coronary events, we are likely to gain power but also obtain a more heterogeneous end point for the analyses. The CVD analyses in ULSAM (I) may also have been affected by the late sampling of DNA at the age of 70, although we investigated this possibility and found no evidence of such bias.

Although the chance of false positive findings cannot be ruled out, we took measures to minimize this in our analyses. Permutation or false discovery rates were applied to control for multiple testing. In FINRISK (III) and in HSDS (II) study samples we also had the opportunity to analyze *USF1* in distinct sample sets to confirm the observed associations, and obtained consistent results with these analyses.

Some of our results were consistent with those from other studies, while other data seemed ambiguous. Two SNPs in our data associated with lipid measurements, atherosclerosis, and with CVD endpoints; *rs2073658* and *rs2516839*. The risk-increasing allele, however, was inconsistent between the study samples; while the common allele of *rs2073658* associated with increased lipid levels in FINRISK (III) men, it lowered them in ULSAM (I) men, protected HSDS (II) men from atherosclerosis, and protected FINRISK (III) women from CVD and mortality. In FINRISK (III), the rare allele of *rs2516839* increased BMI and cholesterol levels of both sexes, but protected HSDS (II) men from atherosclerosis and sudden cardiac death, and decreased BMI of ULSAM (I) men. Rare alleles of both SNPs have also shown both unfavorable and favorable effects in other studies, although the rare allele of *rs2073658* has been protective in most studies (Table 16, Table 17, and Table 18).

The apparent discrepancies in *USF1* studies may be explained, for instance, with the existence of complicated interactions between *USF1* and sex, age, other genes, and environmental factors. Sex-specific gene effects have been reported (Choquette et al. 2007; Lee et al. 2007) and were also present in our study, although they could not be

examined in the ULSAM (I) or HSDS (II) samples, which comprised men only. Studies where any sex-specificity is absent also exist (Huertas-Vazquez et al. 2005; Reiner et al. 2007). These complicating factors could also in part explain the lack of signals at the *USF1* locus in the genome-wide scans for regions contributing to cardiovascular outcomes (Pajukanta et al. 2000; Francke et al. 2001; Broeckel et al. 2002; Harrap et al. 2002; Lange et al. 2002; Hauser et al. 2004; Helgadottir et al. 2004; Wang et al. 2004; Samani et al. 2005; Wang et al. 2005; Engert et al. 2007; Larson et al. 2007; Matarin et al. 2007; Nilsson-Ardnor et al. 2007; O'Donnell et al. 2007).

Notably, in our FINRISK (III) and ULSAM (I) population cohort analyses, the SNP associations were almost exclusively present among CVD or MetS cases. This suggests that a certain pathophysiological state is required for the *USF1* effect to become detectable. Other studies have also suggested this, and especially *USF1*-BMI interactions have been proposed (Putt et al. 2004). Age can also be a significant factor in these analyses; CVD risk and factors contributing to it vary with age, and this can further complicate gene analyses and create discordant results. Especially transcription factors, such as *USF1*, which control several other genes, can affect many biological pathways simultaneously and perhaps contribute to more than one disorder. This hypothesis may be supported by the association of *USF1* SNPs with all-cause mortality in FINRISK (III). Importantly, similar association was also noted in the study by Reiner and colleagues in which the rare allele of *rs2073658* associated with all-cause mortality, just as in FINRISK (III) females, and this association was even stronger in non-CVD mortality (Reiner et al. 2007).

Another complicating factor in our *USF1* study, and in other studies of complex diseases as well, is the number and nature of the SNPs with functional relevance to the disease susceptibility. Many *USF1* studies have only examined a few SNPs instead of comprehensively studying the *USF1* locus for associating variants (Table 16, Table 17, and Table 18). Furthermore, studies on the expression of downstream genes in fat biopsies, have only focused on the initially identified FCHL risk variants (Pajukanta et al. 2004; Naukkarinen et al. 2005). These studies have indicated differential downstream gene expression dependent of *USF1* alleles (Pajukanta et al. 2004; Naukkarinen et al. 2005), and have identified an intronic sequence which binds nuclear protein (Naukkarinen et al. 2005). Although the *USF1* locus does not seem to harbor any common polymorphisms that alter the amino acid sequence of the USF1 protein (Pajukanta et al. 2004; Choquette et al. 2007), SNPs in other critical gene regions could also have a functional consequence. Untranslated variants can, for instance, affect the transcription level of mRNA, splicing patterns and other post-transcriptional modifications of mRNA, or stability or localization of mRNA in the cell. Since the chromosomal region of *USF1* is characterized by high

LD, these functional polymorphisms could reside in the neighboring genes as well (Reiner et al. 2007).

As a ubiquitously expressed transcription factor modulating the expression of multiple genes from lipid, inflammation, and glucose metabolism pathways (Naukkarinen et al. 2005) (Table 8), USF1 could contribute to the features of the traditional CVD risk factors in several ways. Lipid metabolism genes regulated by USF1 include apolipoproteins, such as apolipoprotein C-III (*APOC3*) (Pastier et al. 2002), apolipoprotein A-V (*APOA5*) (Nowak et al. 2005), and apolipoprotein A-II (*APOA2*) (Ribeiro et al. 1999), and other important lipid-related genes such as acetyl-Coenzyme A carboxylase alpha (*ACACA*) (Travers et al. 2001), fatty acid synthase (*FASN*) (Wang and Sul 1997; Casado et al. 1999), hormone sensitive lipase (*LIPE*) (Smih et al. 2002), hepatic lipase (*LIPC*) (Botma et al. 2001), and ATP-binding cassette, sub-family A (ABC1), member 1 (*ABCA1*) (Yang et al. 2002). Furthermore, USF1 is also involved in the regulation of genes which control lipid response to glucose and insulin (Read et al. 1993; Martin et al. 2003a), and inflammation related genes such as C-reactive protein (*CRP*) (Szalai et al. 2005), hepcidin (Bayele et al. 2006), beta-2-microglobulin (*B2m*) (Gobin et al. 2003), and C-4-complement gene (Galibert et al. 1997). Suggestive evidence of an association between *USF1* polymorphisms and inflammation markers, such as plasma CRP and IL6 levels, already exists (Reiner et al. 2007).

Studies on the expression of downstream genes in fat biopsies have already provided evidence of the effect of *USF1* variation at the tissue level (Pajukanta et al. 2004; Naukkarinen et al. 2005). Other studies have demonstrated the involvement of *USF1* in adipocyte lipolysis (Hoffstedt et al. 2005) and in antilipolytic insulin sensitivity (Kantartzis et al. 2007). Future functional studies should include several variants from the *USF1* locus, and should test their functionality in relevant tissues, such as coronary arteries.

Studies with larger sample sizes should also be conducted to confirm our CVD results, which were obtained with relatively modest sample sizes. Future studies should also test for sex-specific effects and possible interactions with BMI or other environmental factors. Finally, upcoming association studies should assess all polymorphisms at the *USF1* locus and neighboring regions, to effectively evaluate the risk increasing or decreasing role of distinct *USF1* alleles.

6 CONCLUDING REMARKS AND FUTURE PROSPECTS

Since the initiation of this study, several publications have described association (or sometimes the lack of it) between the upstream transcription factor 1 (*USF1*) gene and dyslipidemias. Some have sought to relate the gene with metabolic syndrome or type 2 diabetes, and a few have assessed *USF1* variation in relation to cardiovascular disease. The aim of this thesis was to examine how *USF1* locus contributes to CVD risk factors, such as lipid phenotypes and body anthropometrics, at the population level, and to atherosclerosis at the vessel wall level. In addition, sudden cardiac death and incidence of CVD was analyzed. In the process, we also evaluated how whole genome amplification of DNA could improve the amount of samples successfully genotyped in genetic studies like ours, and found that a feasible method exists. Our study provided information supporting a role for *USF1* in the etiology of CVD, and paved the way for future *USF1*-CVD studies.

Cardiovascular diseases are significant contributors to morbidity and mortality worldwide and they have been a subject for intense research. Many, by now conventional, risk factors increasing disease susceptibility are known. In the last few decades, however, research has increasingly focused on the genetic background of these diseases. Despite some success stories, especially concerning rare monogenic forms of CVD, many genes underlying CVD susceptibility remain unknown. These genes act jointly with other genes and environmental factors, in the etiopathogenesis of common, complex forms of CVD.

During the last few years, several improvements in the field of genetics have facilitated the studies of common, complex diseases. Increasing information on single nucleotide polymorphisms (SNPs) and their haplotype-tagging properties in public databases and improved multiplexing in SNP genotyping platforms have enabled several studies with a genome-wide SNP association strategy to emerge. Recent publications have also described the nature of previously overlooked large-scale variation, such as copy number variation, in the human genome. Together, these improvements and new prospects will most likely also lead to many discoveries in the field of CVD genetics.

Identification of genetic risk variants for CVD will not only elucidate the basic biology behind the disease, but will also ultimately provide a means for medicine and pharmacogenomics to develop better treatment and improved prevention.

7 ACKNOWLEDGMENTS

This research was carried out in the Department of Molecular Medicine of the National Public Health Institute, Helsinki, Finland, during the years 2002-2007. The former head of the institute, Jussi Huttunen, and the new director, Pekka Puska, as well as the present Director of the Department of Molecular Medicine, Anu Jalanko, are acknowledged for providing excellent research facilities. This work has been financially supported by the Center of Excellence in Disease Genetics of the Finnish Academy, Biocentrum Helsinki, by the grant numbered 53646 from the Finnish Academy and NIH grants HL-28481 and HL-70150 and the American Heart Association grant 0430180N, Aarne Koskelo Foundation, Emil Aaltonen Foundation, Ernfors Family Foundation, Finnish Foundation for Cardiovascular Research, Jenny and Antti Wihuri Foundation, Knut and Alice Wallenberg Foundation, Medical Research Fund of Tampere University Hospital, Pirkanmaa Regional Fund of the Finnish Cultural Foundation, Sehlander Foundation, Sigrid Juselius Foundation, Swedish Diabetes Association, University of Helsinki, and Yrjö Jahnsson Foundation. They are gratefully acknowledged. The GenomEUtwin project is supported by the European Commission under the program 'Quality of Life and Management of the Living Resources' of the 5th Framework Programme (No. QLG2-CT-2002-01254). I am also indebted to all the subjects for participating in this study.

I want to express my deepest gratitude to my excellent supervisors, Leena Peltonen-Palotie and Markus Perola. I have been privileged to be supervised by these two outstanding scientists with wide expertise and knowledge in the field molecular genetics. I thank them for all the support and responsibility they have given me. Leena's scientific enthusiasm and dedication makes her a great role model for all female scientists. With her inspiring spirit, she has managed to get me to produce my best work. Markus has provided me with the motivation and the skills needed to complete this work. He has also been a tireless source of encouragement.

I would like to thank Juni Palmgren for accepting the role as Opponent in my thesis defense. The reviewers of my thesis, Katariina Kainulainen and Katarina Pelin, are thanked for their encouraging and constructive comments. Katariina Kainulainen and Maija Wessman are thanked for participating in my thesis committee.

My sincere thanks go to our collaborators, for lending their expertise and valuable study samples for this project. Päivi Pajukanta is acknowledged for her help especially at the initiation of this project. Ann-Christine Syvänen, Matti Jauhiainen, Marja-Riitta Taskinen, Björn Zethelius, Christian Berne, and Lars Lannfelt have

contributed to the ULSAM study. Veikko Salomaa, Kari Kuulasmaa, Sangita Kulathinal, Riika Kilpikari, Kaisa Salminen, Olli Saarela, Juha Karvanen, and Zygimantas Cepaitis have been of valuable help in the FINRISK project. The HSDS project has benefited from the expertise of Pekka Karhunen, Terho Lehtimäki, and Erkki Ilveskoski. Pekka is warmly thanked not only for being a skilled expert of his field but also for being one of the most agreeable persons I have had the privilege to work with.

Senior researchers of our laboratory, Anu Jalanko, Ismo Ulmanen, Teppo Varilo, Marjo Kestilä, Iiris Hovatta, Anu Loukola, and especially Janna Saarela, Kaisa Silander, Samuli Ripatti, and Juha Saharinen have helped me in many ways during the past years. Kaisa is particularly thanked for her encouraging advices.

I would like to acknowledge the secretaries Sari Kivikko, Tuija Svahnbäck, Sanna Tossavainen, Helena Knuuttila, Mika Kivimäki, Sari Mustala, and Sami Mustala for their valuable assistance. Jari Raikko, Martti Sariola, Juri Ahokas, Teemu Perheentupa, Juha Knuuttila, and Hannu Turunen are thanked for their help in many computer and database related issues.

Minna Suvela, Arja Tapio, Anne Nyberg, Päivi Tainola, Siv Knaappila, Sisko Lietola, and Liisa Arala are acknowledged for their proficient technical assistance. Pekka Ellonen is thanked for his skillful advices that seem to cover every possible matter in the lab. Mervi Alanne and Outi Törnwall are thanked for their valuable work in the DNA sample managing.

I am forever grateful to my workmates at the office, Mervi, Kirsi, Outi T, and Johannes, for all their help and support during the past years. We have shared brilliant and not so brilliant conversations, numerous laughs, and hilarious nicknames. Your company has unquestionably made the workdays and many social events fun! Tero, Sampo, and Markus, you guys have also been an endless source of witty humor.

Many thanks to all the other special people who have helped me in this project and/or have made our lab such a enjoyable place to work at: Minttu, Pekka, Olli, Marika, Emma P, Emma N, Heidi M, Will, Pia, Juho, Marine, Roxy, Katri, Helena, Liisa T, Annu, Minna, Siv, Sisko, Johanna, Tintti, Annina L, Outi S, Katja, Annina H, Henna, Reetta, Tuuli, Kaisu, Antti, Aimee, Kismat, Suvi, Anu K, Jussi, Pirkka-Pekka, Jaana, and Elli. I have probably forgotten some important names -my apologies for this.

I warmly thank all my friends and relatives for sharing and caring, especially my dear little sisters, Mervi and Maarit, plus Pirjo, Terttu, Eveliina, Ari, and Petri. I thank my parents for all they have taught me. Mervi and Rami, thanks for bringing

sweet little Kasperi into our family. Annikki, Erkki, Vieno, and Aino, I wish you could be here with us today.

Finally, I owe my deepest gratitude to my husband Jouni. Thank you for always believing in me and loving me unconditionally. Without your love and support I would not be where I am today.

Helsinki, 11th of April, 2008

Kati Kristiansson

8 ELECTRONIC DATABASE INFORMATION

AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/)

(Thierry-Mieg et al.)

ArrayExpress (http://www.ebi.ac.uk/arrayexpress/)

ASD, the Alternative Splicing Database

(http://www.ebi.ac.uk/asd/) (Thanaraj et al. 2004; Stamm et al. 2006)

BioGRID (http://www.thebiogrid.org/) (Stark et al. 2006)

Blastn (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al. 1997)

CleanEx (http://www.cleanex.isb-sib.ch/)

dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) (Sherry et al. 2001)

Ensembl (http://www.ensembl.org) (Hubbard et al. 2007)

Entrez Gene (http://www.ncbi.nlm.nih.gov/entrez) (Maglott et al. 2007)

Entrez Nucleotide

(http://www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide&TabCmd=Limits) (Wheeler et al. 2007)

Entrez PopSet (http://www.ncbi.nlm.nih.gov/sites/entrez?db=Popset)

Entrez Protein

(http://www.ncbi.nlm.nih.gov/sites/entrez?db=Protein&TabCmd=Limits) (Wheeler et al. 2007)

Finnish National Cardiovascular Disease Register, statistical database

(http://www.ktl.fi/portal/english/public_health_monitoring___promotion/registers/cardiovascular_diseases/database_overall/) (Mähonen et al. 2000)

Gene Ontology Database (http://www.geneontology.org/) (Ashburner et al. 2000)

GeneCards (http://www.genecards.org) (Rebhan et al. 1997)

GeneLoc (http://bioinfo2.weizmann.ac.il/geneloc) (Rosen et al. 2003)

GEO, Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/)

(Edgar et al. 2002; Barrett et al. 2007)

HapMap (http://www.hapmap.org/) ("The International HapMap Project" 2003)

HGNC (http://www.gene.ucl.ac.uk/nomenclature/)

(Wain et al. 2002; Wain et al. 2004)

HomoloGene (http://www.ncbi.nlm.nih.gov/sites/entrez?db=homologene)

(Wheeler et al. 2007)

Human Protein Reference Database (http://www.hprd.org/)

(Peri et al. 2003; Mishra et al. 2006)

Integrated DNA Technologies (Oligo Analyzer 3.0)

(http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/)

International Classification of Diseases, Ninth Revision (ICD-9)

(http://www.cdc.gov/nchs/about/major/dvs/icd9des.htm)

Oligo Analyzer 3.0 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/)

OMIM (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM)

Primer3 (http://frodo.wi.mit.edu/) (Rozen and Skaletsky 2000)

Repbase Update (http://www.girinst.org/repbase/update/index.html) (Jurka 2000)

SeattleSNPs (http://pga.gs.washington.edu/)

(SeattleSNPs, Variation Discovery Resource)

Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus (http://www.stakes.fi)

Tilastokeskus (http://www.tilastokeskus.fi/) (Tilastokeskus 2007b)

TRANSFAC® Gene Transcription Factor Database (Release 7.0)

(http://www.gene-regulation.com/pub/databases.html) (Matys et al. 2003)

UCSC Genome Browser Database (http://genome.ucsc.edu/)

(Kent et al. 2002; Karolchik et al. 2003)

UniGene (UniGene) (Pontius et al. 2003)

UniProtKB/Swiss-Prot (http://www.ebi.uniprot.org)

(Boeckmann et al. 2003; "The Universal Protein Resource (UniProt)" 2007)

9 REFERENCES

Abifadel, M., Varret, M., Rabes, J. P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derre, A., Villeger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J. M., Luc, G., Moulin, P., Weissenbach, J., Prat, A., Krempf, M., Junien, C., Seidah, N. G., and Boileau, C. (2003). "Mutations in PCSK9 cause autosomal dominant hypercholesterolemia." *Nat Genet* 34(2): 154-156.

Achar, S. A., Kundu, S., and Norcross, W. A. (2005). "Diagnosis of Acute Coronary Syndrome." *Am Fam Physician* 72(1): 119-126.

Alanne, M., Salomaa, V., Saarela, J., Peltonen, L., and Perola, M. (2004). "DNA extraction yield is associated with several phenotypic characteristics: results from two large population surveys." *J Thromb Haemost* 2(11): 2069-2071.

Alberti, K. G., Zimmet, P., and Shaw, J. (2005). "The metabolic syndrome--a new worldwide definition." *Lancet* 366(9491): 1059-1062.

Alberti, K. G., and Zimmet, P. Z. (1998). "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation." *Diabet Med* 15(7): 539-553.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25(17): 3389-3402.

Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J., Lane, C. R., Schaffner, S. F., Bolk, S., Brewer, C., Tuomi, T., Gaudet, D., Hudson, T. J., Daly, M., Groop, L., and Lander, E. S. (2000). "The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes." *Nat Genet* 26(1): 76-80.

Andrews, G. K. (2000). "Regulation of metallothionein gene expression by oxidative stress and metal ions." *Biochem Pharmacol* 59(1): 95-104.

Arnett, D. K., Baird, A. E., Barkley, R. A., Basson, C. T., Boerwinkle, E., Ganesh, S. K., Herrington, D. M., Hong, Y., Jaquish, C., McDermott, D. A., and O'Donnell, C. J. (2007). "Relevance of genetics and genomics for prevention and treatment of cardiovascular disease: a scientific statement from the American Heart Association Council on Epidemiology and Prevention, the Stroke Council, and the Functional Genomics and Translational Biology Interdisciplinary Working Group." *Circulation* 115(22): 2878-2901.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nat Genet* 25(1): 25-29.

Auro, K., Kristiansson, K., Zethelius, B., Berne, C., Lannfelt, L., Taskinen, M. R., Jauhiainen, M., Perola, M., Peltonen, L., and Syvanen, A. C. (2007). "USF1 gene variants contribute to metabolic traits in men in a longitudinal 32-year follow-up study." *Diabetologia*.

Austin, M. A., McKnight, B., Edwards, K. L., Bradley, C. M., McNeely, M. J., Psaty, B. M., Brunzell, J. D., and Motulsky, A. G. (2000). "Cardiovascular disease mortality in familial forms of hypertriglyceridemia: A 20-year prospective study." *Circulation* 101(24): 2777-2782.

Babinska, A., Azari, B. M., Salifu, M. O., Liu, R., Jiang, X. C., Sobocka, M. B., Boo, D., Al Khoury, G., Deitch, J. S., Marmur, J. D., Ehrlich, Y. H., and Kornecki, E. (2007). "The F11 receptor (F11R/JAM-A) in atherothrombosis: overexpression of F11R in atherosclerotic plaques." *Thromb Haemost* 97(2): 272-281.

Babinska, A., Kedees, M. H., Athar, H., Ahmed, T., Batuman, O., Ehrlich, Y. H., Hussain, M. M., and Kornecki, E. (2002). "F11-receptor (F11R/JAM) mediates platelet adhesion to endothelial cells: role in inflammatory thrombosis." *Thromb Haemost* 88(5): 843-850.

Bak, S., Gaist, D., Sindrup, S. H., Skytthe, A., and Christensen, K. (2002). "Genetic liability in stroke: a long-term follow-up study of Danish twins." *Stroke* 33(3): 769-774.

Barlow, W. E. (1994). "Robust variance estimation for the case-cohort design." *Biometrics* 50(4): 1064-1072.

Barrett-Connor, E. (1997). "Sex differences in coronary heart disease. Why are women so superior? The 1995 Ancel Keys Lecture." *Circulation* 95(1): 252-264.

Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). "Haploview: analysis and visualization of LD and haplotype maps." *Bioinformatics* 21(2): 263-265.

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). "NCBI GEO: mining tens of millions of expression profiles--database and tools update." *Nucleic Acids Res* 35(Database issue): D760-765.

Bayele, H. K., McArdle, H., and Srai, S. K. (2006). "Cis and trans regulation of hepcidin expression by upstream stimulatory factor." *Blood* 108(13): 4237-4245.

Bayoumi, R. A., Al-Yahyaee, S. A., Albarwani, S. A., Rizvi, S. G., Al-Hadabi, S., Al-Ubaidi, F. F., Al-Hinai, A. T., Al-Kindi, M. N., Adnan, H. T., Al-Barwany, H. S., Comuzzie, A. G., Cai, G., Lopez-Alvarenga, J. C., and Hassan, M. O. (2007). "Heritability of determinants of the metabolic syndrome among healthy Arabs of the Oman family study." *Obesity (Silver Spring)* 15(3): 551-556.

Berenson, G. S., Srinivasan, S. R., Bao, W., Newman, W. P., 3rd, Tracy, R. E., and Wattigney, W. A. (1998). "Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study." *N Engl J Med* 338(23): 1650-1656.

Berge, K. E., Tian, H., Graf, G. A., Yu, L., Grishin, N. V., Schultz, J., Kwiterovich, P., Shan, B., Barnes, R., and Hobbs, H. H. (2000). "Accumulation of dietary cholesterol in

sitosterolemia caused by mutations in adjacent ABC transporters." *Science* 290(5497): 1771-1775.

Bertina, R. M., Koeleman, B. P., Koster, T., Rosendaal, F. R., Dirven, R. J., de Ronde, H., van der Velden, P. A., and Reitsma, P. H. (1994). "Mutation in blood coagulation factor V associated with resistance to activated protein C." *Nature* 369(6475): 64-67.

Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., Kaminski, W. E., Hahmann, H. W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K. J., and Schmitz, G. (1999). "The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease." *Nat Genet* 22(4): 347-351.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003). "The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003." *Nucleic Acids Res* 31(1): 365-370.

Boekholdt, S. M., Bijsterveld, N. R., Moons, A. H., Levi, M., Buller, H. R., and Peters, R. J. (2001). "Genetic variation in coagulation and fibrinolytic proteins and their relation with acute myocardial infarction: a systematic review." *Circulation* 104(25): 3063-3068.

Boekholdt, S. M., Sacks, F. M., Jukema, J. W., Shepherd, J., Freeman, D. J., McMahon, A. D., Cambien, F., Nicaud, V., de Grooth, G. J., Talmud, P. J., Humphries, S. E., Miller, G. J., Eiriksdottir, G., Gudnason, V., Kauma, H., Kakko, S., Savolainen, M. J., Arca, M., Montali, A., Liu, S., Lanz, H. J., Zwinderman, A. H., Kuivenhoven, J. A., and Kastelein, J. J. (2005). "Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects." *Circulation* 111(3): 278-287.

Botma, G. J., Verhoeven, A. J., and Jansen, H. (2001). "Hepatic lipase promoter activity is reduced by the C-480T and G-216A substitutions present in the common LIPC gene variant, and is increased by Upstream Stimulatory Factor." *Atherosclerosis* 154(3): 625-632.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." *Am J Hum Genet* 32(3): 314-331.

Broeckel, U., Hengstenberg, C., Mayer, B., Holmer, S., Martin, L. J., Comuzzie, A. G., Blangero, J., Nurnberg, P., Reis, A., Riegger, G. A., Jacob, H. J., and Schunkert, H. (2002). "A comprehensive linkage analysis for myocardial infarction and its related risk factors." *Nat Genet* 30(2): 210-214.

Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L. H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J. A., Molhuizen, H. O., Loubser, O., Ouelette, B. F., Fichter, K., Ashbourne-Excoffon, K. J., Sensen, C. W., Scherer, S., Mott, S., Denis, M., Martindale, D., Frohlich, J., Morgan, K., Koop, B., Pimstone, S., Kastelein, J. J., Genest, J., Jr., and Hayden, M. R. (1999). "Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency." *Nat Genet* 22(4): 336-345.

Brown, W. M., Beck, S. R., Lange, E. M., Davis, C. C., Kay, C. M., Langefeld, C. D., and Rich, S. S. (2003). "Age-stratified heritability estimation in the Framingham Heart Study families." *BMC Genet* 4 Suppl 1: S32.

Burton, P. R., Tobin, M. D., and Hopper, J. L. (2005). "Key concepts in genetic epidemiology." *Lancet* 366(9489): 941-951.

Burzotta, F., Paciaroni, K., De Stefano, V., Crea, F., Maseri, A., Leone, G., and Andreotti, F. (2004). "G20210A prothrombin gene polymorphism and coronary ischaemic syndromes: a phenotype-specific meta-analysis of 12 034 subjects." *Heart* 90(1): 82-86.

Cambien, F., and Tiret, L. (2007). "Genetics of cardiovascular diseases: from single mutations to the whole genome." *Circulation* 116(15): 1714-1724.

Cardon, L. R., and Bell, J. I. (2001). "Association study designs for complex diseases." *Nat Rev Genet* 2(2): 91-99.

Cardon, L. R., and Palmer, L. J. (2003). "Population stratification and spurious allelic association." *Lancet* 361(9357): 598-604.

Carlson, C. S., Eberle, M. A., Kruglyak, L., and Nickerson, D. A. (2004). "Mapping complex disease loci in whole-genome association studies." *Nature* 429(6990): 446-452.

Casado, M., Vallet, V. S., Kahn, A., and Vaulont, S. (1999). "Essential role in vivo of upstream stimulatory factors for a normal dietary response of the fatty acid synthase gene in the liver." *J Biol Chem* 274(4): 2009-2013.

Casas, J. P., Cooper, J., Miller, G. J., Hingorani, A. D., and Humphries, S. E. (2006). "Investigating the genetic determinants of cardiovascular disease using candidate genes and meta-analysis of association studies." *Ann Hum Genet* 70(Pt 2): 145-169.

Cassidy-Bushrow, A. E., Bielak, L. F., Sheedy, P. F., 2nd, Turner, S. T., Kullo, I. J., Lin, X., and Peyser, P. A. (2007). "Coronary artery calcification progression is heritable." *Circulation* 116(1): 25-31.

Caulfield, M., Lavender, P., Farrall, M., Munroe, P., Lawson, M., Turner, P., and Clark, A. J. (1994). "Linkage of the angiotensinogen gene to essential hypertension." *N Engl J Med* 330(23): 1629-1633.

Cheung, V. G., and Nelson, S. F. (1996). "Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA." *Proc Natl Acad Sci U S A* 93(25): 14676-14679.

Chiodini, B. D., Barlera, S., Franzosi, M. G., Beceiro, V. L., Introna, M., and Tognoni, G. (2003). "APO B gene polymorphisms and coronary artery disease: a meta-analysis." *Atherosclerosis* 167(2): 355-366.

Chiodini, B. D., and Lewis, C. M. (2003). "Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q." *Arterioscler Thromb Vasc Biol* 23(10): 1863-1868.

Choquette, A. C., Bouchard, L., Houde, A., Bouchard, C., Perusse, L., and Vohl, M. C. (2007). "Associations between USF1 gene variants and cardiovascular risk factors in the Quebec Family Study." *Clin Genet* 71(3): 245-253.

Coon, H., Xin, Y., Hopkins, P. N., Cawthon, R. M., Hasstedt, S. J., and Hunt, S. C. (2005). "Upstream stimulatory factor 1 associated with familial combined hyperlipidemia, LDL cholesterol, and triglycerides." *Hum Genet* 117(5): 444-451.

Cordell, H. J., and Clayton, D. G. (2005). "Genetic association studies." *Lancet* 366(9491): 1121-1131.

Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). "General nature of the genetic code for proteins." *Nature* 192: 1227-1232.

Cushwa, W. T., and Medrano, J. F. (1993). "Effects of blood storage time and temperature on DNA yield and quality." *Biotechniques* 14(2): 204-207.

Dale, A. C., Nilsen, T. I., Vatten, L., Midthjell, K., and Wiseth, R. (2007). "Diabetes mellitus and risk of fatal ischaemic heart disease by gender: 18 years follow-up of 74 914 individuals in the HUNT 1 Study." *Eur Heart J*.

Danesh, J., Saracci, R., Berglund, G., Feskens, E., Overvad, K., Panico, S., Thompson, S., Fournier, A., Clavel-Chapelon, F., Canonico, M., Kaaks, R., Linseisen, J., Boeing, H., Pischon, T., Weikert, C., Olsen, A., Tjonneland, A., Johnsen, S. P., Jensen, M. K., Quiros, J. R., Svatetz, C. A., Perez, M. J., Larranaga, N., Sanchez, C. N., Iribas, C. M., Bingham, S., Khaw, K. T., Wareham, N., Key, T., Roddam, A., Trichopoulou, A., Benetou, V., Trichopoulos, D., Masala, G., Sieri, S., Tumino, R., Sacerdote, C., Mattiello, A., Verschuren, W. M., Bueno-de-Mesquita, H. B., Grobbee, D. E., van der Schouw, Y. T., Melander, O., Hallmans, G., Wennberg, P., Lund, E., Kumle, M., Skeie, G., Ferrari, P., Slimani, N., Norat, T., and Riboli, E. (2007). "EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries." *Eur J Epidemiol* 22(2): 129-141.

Dawn Teare, M., and Barrett, J. H. (2005). "Genetic linkage studies." *Lancet* 366(9490): 1036-1044.

Dean, F. B., Nelson, J. R., Giesler, T. L., and Lasken, R. S. (2001). "Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification." *Genome Res* 11(6): 1095-1099.

Dietmaier, W., Hartmann, A., Wallinger, S., Heinmoller, E., Kerner, T., Endl, E., Jauch, K. W., Hofstadter, F., and Ruschoff, J. (1999). "Multiple mutation analyses in single tumor cells with improved whole genome amplification." *Am J Pathol* 154(1): 83-95.

Eckel, R. H., Grundy, S. M., and Zimmet, P. Z. (2005). "The metabolic syndrome." *Lancet* 365(9468): 1415-1428.

Edgar, R., Domrachev, M., and Lash, A. E. (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository." *Nucleic Acids Res* 30(1): 207-210.

Elbein, S. C., Hoffman, M. D., Teng, K., Leppert, M. F., and Hasstedt, S. J. (1999). "A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians." *Diabetes* 48(5): 1175-1182.

Engert, J. C., Lemire, M., Faith, J., Brisson, D., Fujiwara, T. M., Roslin, N. M., Brewer, C. G., Montpetit, A., Darmond-Zwaig, C., Renaud, Y., Dore, C., Bailey, S. D., Verner, A., Tremblay, G., St-Pierre, J., Betard, C., Platko, J., Rioux, J. D., Morgan, K., Hudson, T. J., and Gaudet, D. (2007). "Identification of a chromosome 8p locus for early-onset coronary heart disease in a French Canadian population." *Eur J Hum Genet*.

Evans, A., Salomaa, V., Kulathinal, S., Asplund, K., Cambien, F., Ferrario, M., Perola, M., Peltonen, L., Shields, D., Tunstall-Pedoe, H., and Kuulasmaa, K. (2005). "MORGAM (an international pooling of cardiovascular cohorts)." *Int J Epidemiol* 34(1): 21-27.

Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G., and Burley, S. K. (1994). "Structure and function of the b/HLH/Z domain of USF." *Embo J* 13(1): 180-189.

Feuk, L., Carson, A. R., and Scherer, S. W. (2006). "Structural variation in the human genome." *Nat Rev Genet* 7(2): 85-97.

Fischer, M., Broeckel, U., Holmer, S., Baessler, A., Hengstenberg, C., Mayer, B., Erdmann, J., Klein, G., Riegger, G., Jacob, H. J., and Schunkert, H. (2005). "Distinct heritable patterns of angiographic coronary artery disease in families with myocardial infarction." *Circulation* 111(7): 855-862.

Flossmann, E., Schulz, U. G., and Rothwell, P. M. (2004). "Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke." *Stroke* 35(1): 212-227.

Fong, I. W. (2002). "Infections and their role in atherosclerotic vascular disease." *J Am Dent Assoc* 133 Suppl: 7S-13S.

Fox, C. S., Polak, J. F., Chazaro, I., Cupples, A., Wolf, P. A., D'Agostino, R. A., and O'Donnell, C. J. (2003). "Genetic and environmental contributions to atherosclerosis phenotypes in men and women: heritability of carotid intima-media thickness in the Framingham Heart Study." *Stroke* 34(2): 397-401.

Francke, S., Manraj, M., Lacquemant, C., Lecoeur, C., Lepretre, F., Passa, P., Hebe, A., Corset, L., Yan, S. L., Lahmidi, S., Jankee, S., Gunness, T. K., Ramjuttun, U. S., Balgobin, V., Dina, C., and Froguel, P. (2001). "A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27." *Hum Mol Genet* 10(24): 2751-2765.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A. M., Ness, A. R., Ebrahim, S., Lawlor, D. A., Ring, S. M., Ben-Shlomo, Y., Järvelin, M. R., Sovio, U., Bennett, A. J., Melzer, D., Ferrucci, L., Loos, R. J., Barroso, I., Wareham, N. J., Karpe, F., Owen, K. R., Cardon, L. R., Walker, M., Hitman, G. A., Palmer, C. N., Doney, A. S., Morris, A. D., Smith, G. D., Hattersley, A. T., and McCarthy, M. I. (2007). "A common variant in

the FTO gene is associated with body mass index and predisposes to childhood and adult obesity." *Science* 316(5826): 889-894.

Frizzell, J. P. (2005). "Acute stroke: pathophysiology, diagnosis, and treatment." *AACN Clin Issues* 16(4): 421-440; quiz 597-428.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., and Altshuler, D. (2002). "The structure of haplotype blocks in the human genome." *Science* 296(5576): 2225-2229.

Galibert, M. D., Boucontet, L., Goding, C. R., and Meo, T. (1997). "Recognition of the E-C4 element from the C4 complement gene promoter by the upstream stimulatory factor-1 transcription factor." *J Immunol* 159(12): 6176-6183.

Genest, J. J., Jr., Martin-Munley, S. S., McNamara, J. R., Ordovas, J. M., Jenner, J., Myers, R. H., Silberman, S. R., Wilson, P. W., Salem, D. N., and Schaefer, E. J. (1992). "Familial lipoprotein disorders in patients with premature coronary artery disease." *Circulation* 85(6): 2025-2033.

Georges, J. L., Loukaci, V., Poirier, O., Evans, A., Luc, G., Arveiler, D., Ruidavets, J. B., Cambien, F., and Tiret, L. (2001). "Interleukin-6 gene polymorphisms and susceptibility to myocardial infarction: the ECTIM study. Etude Cas-Temoin de l'Infarctus du Myocarde." *J Mol Med* 79(5-6): 300-305.

Gibson, F., Hercberg, S., and Froguel, P. (2005). "Common polymorphisms in the USF1 gene are not associated with type 2 diabetes in French Caucasians." *Diabetes* 54(10): 3040-3042.

Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). "Finding genes that underlie complex traits." *Science* 298(5602): 2345-2349.

Gobin, S. J., Biesta, P., and Van den Elsen, P. J. (2003). "Regulation of human beta 2-microglobulin transactivation in hematopoietic cells." *Blood* 101(8): 3058-3064.

Goldstein, J. L., and Brown, M. S. (1979). "The LDL receptor locus and the genetics of familial hypercholesterolemia." *Annu Rev Genet* 13: 259-289.

Goldstein, J. L., Hazzard, W. R., Schrott, H. G., Bierman, E. L., and Motulsky, A. G. (1973a). "Hyperlipidemia in coronary heart disease. I. Lipid levels in 500 survivors of myocardial infarction." *J Clin Invest* 52(7): 1533-1543.

Goldstein, J. L., Schrott, H. G., Hazzard, W. R., Bierman, E. L., and Motulsky, A. G. (1973b). "Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia." *J Clin Invest* 52(7): 1544-1568.

Goldstein, L. B., Adams, R., Alberts, M. J., Appel, L. J., Brass, L. M., Bushnell, C. D., Culebras, A., Degraba, T. J., Gorelick, P. B., Guyton, J. R., Hart, R. G., Howard, G., Kelly-Hayes, M., Nixon, J. V., and Sacco, R. L. (2006). "Primary prevention of ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council: cosponsored by the Atherosclerotic Peripheral Vascular Disease

Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; and the Quality of Care and Outcomes Research Interdisciplinary Working Group: the American Academy of Neurology affirms the value of this guideline." *Stroke* 37(6): 1583-1633.

Grant, S. F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., Styrkarsdottir, U., Magnusson, K. P., Walters, G. B., Palsdottir, E., Jonsdottir, T., Gudmundsdottir, T., Gylfason, A., Saemundsdottir, J., Wilensky, R. L., Reilly, M. P., Rader, D. J., Bagger, Y., Christiansen, C., Gudnason, V., Sigurdsson, G., Thorsteinsdottir, U., Gulcher, J. R., Kong, A., and Stefansson, K. (2006). "Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes." *Nat Genet* 38(3): 320-323.

Grech, E. D., and Ramsdale, D. R. (2003). "Acute coronary syndrome: unstable angina and non-ST segment elevation myocardial infarction." *Bmj* 326: 1259-1261.

Gregor, P. D., Sawadogo, M., and Roeder, R. G. (1990). "The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer." *Genes Dev* 4(10): 1730-1740.

Gretarsdottir, S., Thorleifsson, G., Reynisdottir, S. T., Manolescu, A., Jonsdottir, S., Jonsdottir, T., Gudmundsdottir, T., Bjarnadottir, S. M., Einarsson, O. B., Gudjonsdottir, H. M., Hawkins, M., Gudmundsson, G., Gudmundsdottir, H., Andrason, H., Gudmundsdottir, A. S., Sigurdardottir, M., Chou, T. T., Nahmias, J., Goss, S., Sveinbjornsdottir, S., Valdimarsson, E. M., Jakobsson, F., Agnarsson, U., Gudnason, V., Thorgeirsson, G., Fingerle, J., Gurney, M., Gudbjartsson, D., Frigge, M. L., Kong, A., Stefansson, K., and Gulcher, J. R. (2003). "The gene encoding phosphodiesterase 4D confers risk of ischemic stroke." *Nat Genet* 35(2): 131-138.

Hanson, R. L., Ehm, M. G., Pettitt, D. J., Prochazka, M., Thompson, D. B., Timberlake, D., Foroud, T., Kobes, S., Baier, L., Burns, D. K., Almasy, L., Blangero, J., Garvey, W. T., Bennett, P. H., and Knowler, W. C. (1998). "An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians." *Am J Hum Genet* 63(4): 1130-1138.

Hanson, R. L., Imperatore, G., Narayan, K. M., Roumain, J., Fagot-Campagna, A., Pettitt, D. J., Bennett, P. H., and Knowler, W. C. (2001). "Family and genetic studies of indices of insulin sensitivity and insulin secretion in Pima Indians." *Diabetes Metab Res Rev* 17(4): 296-303.

Harrap, S. B., Zammit, K. S., Wong, Z. Y., Williams, F. M., Bahlo, M., Tonkin, A. M., and Anderson, S. T. (2002). "Genome-wide linkage analysis of the acute coronary syndrome suggests a locus on chromosome 2." *Arterioscler Thromb Vasc Biol* 22(5): 874-878.

Hauser, E. R., Crossman, D. C., Granger, C. B., Haines, J. L., Jones, C. J., Mooser, V., McAdam, B., Winkelmann, B. R., Wiseman, A. H., Muhlestein, J. B., Bartel, A. G., Dennis, C. A., Dowdy, E., Estabrooks, S., Eggleston, K., Francis, S., Roche, K., Clevenger, P. W., Huang, L., Pedersen, B., Shah, S., Schmidt, S., Haynes, C., West, S., Asper, D., Booze, M., Sharma, S., Sundseth, S., Middleton, L., Roses, A. D., Hauser, M. A., Vance, J. M.,

Pericak-Vance, M. A., and Kraus, W. E. (2004). "A genomewide scan for early-onset coronary artery disease in 438 families: the GENECARD Study." *Am J Hum Genet* 75(3): 436-447.

Hawkins, T. L., Detter, J. C., and Richardson, P. M. (2002). "Whole genome amplification--applications and advances." *Curr Opin Biotechnol* 13(1): 65-67.

Helgadottir, A., Manolescu, A., Thorleifsson, G., Gretarsdottir, S., Jonsdottir, H., Thorsteinsdottir, U., Samani, N. J., Gudmundsson, G., Grant, S. F., Thorgeirsson, G., Sveinbjornsdottir, S., Valdimarsson, E. M., Matthiasson, S. E., Johannsson, H., Gudmundsdottir, O., Gurney, M. E., Sainz, J., Thorhallsdottir, M., Andresdottir, M., Frigge, M. L., Topol, E. J., Kong, A., Gudnason, V., Hakonarson, H., Gulcher, J. R., and Stefansson, K. (2004). "The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke." *Nat Genet* 36(3): 233-239.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A., Masson, G., Gudbjartsson, D. F., Magnusson, K. P., Andersen, K., Levey, A. I., Backman, V. M., Matthiasdottir, S., Jonsdottir, T., Palsson, S., Einarsdottir, H., Gunnarsdottir, S., Gylfason, A., Vaccarino, V., Hooper, W. C., Reilly, M. P., Granger, C. B., Austin, H., Rader, D. J., Shah, S. H., Quyyumi, A. A., Gulcher, J. R., Thorgeirsson, G., Thorsteinsdottir, U., Kong, A., and Stefansson, K. (2007). "A common variant on chromosome 9p21 affects the risk of myocardial infarction." *Science* 316(5830): 1491-1493.

Heller, D. A., de Faire, U., Pedersen, N. L., Dahlen, G., and McClearn, G. E. (1993). "Genetic and environmental influences on serum lipid levels in twins." *N Engl J Med* 328(16): 1150-1156.

Herbert, A., Gerry, N. P., McQueen, M. B., Heid, I. M., Pfeufer, A., Illig, T., Wichmann, H. E., Meitinger, T., Hunter, D., Hu, F. B., Colditz, G., Hinney, A., Hebebrand, J., Koberwitz, K., Zhu, X., Cooper, R., Ardlie, K., Lyon, H., Hirschhorn, J. N., Laird, N. M., Lenburg, M. E., Lange, C., and Christman, M. F. (2006). "A common genetic variant is associated with adult and childhood obesity." *Science* 312(5771): 279-283.

Heutink, P., and Oostra, B. A. (2002). "Gene finding in genetically isolated populations." *Hum Mol Genet* 11(20): 2507-2515.

Hirschhorn, J. N., and Daly, M. J. (2005). "Genome-wide association studies for common diseases and complex traits." *Nat Rev Genet* 6(2): 95-108.

Hoffstedt, J., Ryden, M., Wahrenberg, H., van Harmelen, V., and Arner, P. (2005). "Upstream transcription factor-1 gene polymorphism is associated with increased adipocyte lipolysis." *J Clin Endocrinol Metab* 90(9): 5356-5360.

Hopkins, P. N., Heiss, G., Ellison, R. C., Province, M. A., Pankow, J. S., Eckfeldt, J. H., and Hunt, S. C. (2003). "Coronary artery disease risk in familial combined hyperlipidemia and familial hypertriglyceridemia: a case-control comparison from the National Heart, Lung, and Blood Institute Family Heart Study." *Circulation* 108(5): 519-523.

Hsueh, W. C., St Jean, P. L., Mitchell, B. D., Pollin, T. I., Knowler, W. C., Ehm, M. G., Bell, C. J., Sakul, H., Wagner, M. J., Burns, D. K., and Shuldiner, A. R. (2003). "Genome-wide and

fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24." *Diabetes* 52(2): 550-557.

Hubbard, T. J., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., and Birney, E. (2007). "Ensembl 2007." *Nucleic Acids Res* 35(Database issue): D610-617.

Huertas-Vazquez, A., Aguilar-Salinas, C., Lusis, A. J., Cantor, R. M., Canizales-Quinteros, S., Lee, J. C., Mariana-Nunez, L., Riba-Ramirez, R. M., Jokiaho, A., Tusie-Luna, T., and Pajukanta, P. (2005). "Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1." *Arterioscler Thromb Vasc Biol* 25(9): 1985-1991.

Huxley, R., Barzi, F., and Woodward, M. (2006). "Excess risk of fatal coronary heart disease associated with diabetes in men and women: meta-analysis of 37 prospective cohort studies." *Bmj* 332(7533): 73-78.

Ingelsson, E., Larson, M. G., Vasan, R. S., O'Donnell, C. J., Yin, X., Hirschhorn, J. N., Newton-Cheh, C., Drake, J. A., Musone, S. L., Heard-Costa, N. L., Benjamin, E. J., Levy, D., Atwood, L. D., Wang, T. J., and Kathiresan, S. (2007). "Heritability, linkage, and genetic associations of exercise treadmill test responses." *Circulation* 115(23): 2917-2924.

Innerarity, T. L., Weisgraber, K. H., Arnold, K. S., Mahley, R. W., Krauss, R. M., Vega, G. L., and Grundy, S. M. (1987). "Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding." *Proc Natl Acad Sci U S A* 84(19): 6919-6923.

"The International HapMap Project." (2003). *Nature* 426(6968): 789-796.

International Human Genome Sequencing Consortium (2004). "Finishing the euchromatic sequence of the human genome." *Nature* 431(7011): 931-945.

Isola, J., DeVries, S., Chu, L., Ghazvini, S., and Waldman, F. (1994). "Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples." *Am J Pathol* 145(6): 1301-1308.

Isomaa, B., Almgren, P., Tuomi, T., Forsen, B., Lahti, K., Nissen, M., Taskinen, M. R., and Groop, L. (2001). "Cardiovascular morbidity and mortality associated with the metabolic syndrome." *Diabetes Care* 24(4): 683-689.

Jobling, M. A., Hurles, M. E., and Tyler-Smith, C. (2004). *Human Evolutionary Genetics: Origins, Peoples & Disease*, Garland Science.

Johnson, G. C., Esposito, L., Barratt, B. J., Smith, A. N., Heward, J., Di Genova, G., Ueda, H., Cordell, H. J., Eaves, I. A., Dudbridge, F., Twells, R. C., Payne, F., Hughes, W., Nutland, S., Stevens, H., Carr, P., Tuomilehto-Wolf, E., Tuomilehto, J., Gough, S. C., Clayton, D. G., and Todd, J. A. (2001). "Haplotype tagging for the identification of common disease genes." *Nat Genet* 29(2): 233-237.

Jorde, L. B. (2000). "Linkage disequilibrium and the search for complex disease genes." *Genome Res* 10(10): 1435-1444.

Jousilahti, P., Puska, P., Vartiainen, E., Pekkanen, J., and Tuomilehto, J. (1996). "Parental history of premature coronary heart disease: an independent risk factor of myocardial infarction." *J Clin Epidemiol* 49(5): 497-503.

Jurka, J. (2000). "Repbase update: a database and an electronic journal of repetitive elements." *Trends Genet* 16(9): 418-420.

Kantartzis, K., Fritsche, A., Machicao, F., Stumvoll, M., Machann, J., Schick, F., Haring, H. U., and Stefan, N. (2007). "Upstream transcription factor 1 gene polymorphisms are associated with high antilipolytic insulin sensitivity and show gene-gene interactions." *Journal of Molecular Medicine* 85(1): 55-61.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., Weber, R. J., Haussler, D., and Kent, W. J. (2003). "The UCSC Genome Browser Database." *Nucleic Acids Res* 31(1): 51-54.

Kattainen, A., Salomaa, V., Härkänen, T., Jula, A., Kaaja, R., Kesäniemi, Y. A., Kähönen, M., Moilanen, L., Nieminen, M. S., Aromaa, A., and Reunanen, A. (2006). "Coronary heart disease: from a disease of middle-aged men in the late 1970s to a disease of elderly women in the 2000s." *Eur Heart J* 27(3): 296-301.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). "The human genome browser at UCSC." *Genome Res* 12(6): 996-1006.

Keskitalo, K., Knaapila, A., Kallela, M., Palotie, A., Wessman, M., Sammalisto, S., Peltonen, L., Tuorila, H., and Perola, M. (2007). "Sweet taste preferences are partly genetically determined: identification of a trait locus on chromosome 16." *Am J Clin Nutr* 86(1): 55-63.

Khorana, H. G. (1968). "Synthesis in the study of nucleic acids. The Fourth Jubilee Lecture." *Biochem J* 109(5): 709-725.

Kiely, D. K., Wolf, P. A., Cupples, L. A., Beiser, A. S., and Myers, R. H. (1993). "Familial aggregation of stroke. The Framingham Study." *Stroke* 24(9): 1366-1371.

Kim, R. J., and Becker, R. C. (2003). "Association between factor V Leiden, prothrombin G20210A, and methylenetetrahydrofolate reductase C677T mutations and events of the arterial circulatory system: a meta-analysis of published studies." *Am Heart J* 146(6): 948-957.

Klerk, M., Verhoef, P., Clarke, R., Blom, H. J., Kok, F. J., and Schouten, E. G. (2002). "MTHFR 677C-->T polymorphism and risk of coronary heart disease: a meta-analysis." *Jama* 288(16): 2023-2031.

Komulainen, K., Alanne, M., Auro, K., Kilpikari, R., Pajukanta, P., Saarela, J., Ellonen, P., Salminen, K., Kulathinal, S., Kuulasmaa, K., Silander, K., Salomaa, V., Perola, M., and Peltonen, L. (2006). "Risk alleles of USF1 gene predict cardiovascular disease of women in two prospective studies." *PLoS Genet* 2(5): e69.

Kristiansson, K. (2007). *USF1*-geenin polymorfismien yhteys sydän- ja verisuonitautien ilmaantuvuuteen suomalaisessa seuranta-aineistossa. *Biotieteellinen tiedekunta, Bio- ja ympäristötieteiden laitos, Perinnöllisyystiede*. Helsinki, University of Helsinki**:** 99.

Kronzon, I., and Tunick, P. A. (2006). "Aortic Atherosclerotic Disease and Stroke." *Circulation* 114: 63-75.

Kuulasmaa, K., Tunstall-Pedoe, H., Dobson, A., Fortmann, S., Sans, S., Tolonen, H., Evans, A., Ferrario, M., and Tuomilehto, J. (2000). "Estimation of contribution of changes in classic risk factors to trends in coronary-event rates across the WHO MONICA Project populations." *Lancet* 355(9205): 675-687.

Lage, J. M., Leamon, J. H., Pejovic, T., Hamann, S., Lacey, M., Dillon, D., Segraves, R., Vossbrinck, B., Gonzalez, A., Pinkel, D., Albertson, D. G., Costa, J., and Lizardi, P. M. (2003). "Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH." *Genome Res* 13(2): 294-307.

Lakka, H. M., Laaksonen, D. E., Lakka, T. A., Niskanen, L. K., Kumpusalo, E., Tuomilehto, J., and Salonen, J. T. (2002). "The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men." *Jama* 288(21): 2709-2716.

Landegren, U., Nilsson, M., and Kwok, P. Y. (1998). "Reading bits of genetic information: methods for single-nucleotide polymorphism analysis." *Genome Res* 8(8): 769-776.

Lander, E., and Kruglyak, L. (1995). "Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results." *Nat Genet* 11(3): 241-247.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E.,

Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. J. (2001). "Initial sequencing and analysis of the human genome." *Nature* 409(6822): 860-921.

Lander, E. S., and Schork, N. J. (1994). "Genetic dissection of complex traits." *Science* 265(5181): 2037-2048.

Lange, L. A., Lange, E. M., Bielak, L. F., Langefeld, C. D., Kardia, S. L., Royston, P., Turner, S. T., Sheedy, P. F., 2nd, Boerwinkle, E., and Peyser, P. A. (2002). "Autosomal genome-wide scan for coronary artery calcification loci in sibships at high risk for hypertension." *Arterioscler Thromb Vasc Biol* 22(3): 418-423.

Larson, M. G., Atwood, L. D., Benjamin, E. J., Cupples, L. A., D'Agostino, R. B., Sr., Fox, C. S., Govindaraju, D. R., Guo, C. Y., Heard-Costa, N. L., Hwang, S. J., Murabito, J. M., Newton-Cheh, C., O'Donnell, C. J., Seshadri, S., Vasan, R. S., Wang, T. J., Wolf, P. A., and Levy, D. (2007). "Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes." *BMC Med Genet* 8 Suppl 1: S5.

Lee, J. C., Weissglas-Volkov, D., Kyttälä, M., Sinsheimer, J. S., Jokiaho, A., de Bruin, T. W., Lusis, A. J., Brennan, M. L., van Greevenbroek, M. M., van der Kallen, C. J., Hazen, S. L., and Pajukanta, P. (2007). "USF1 Contributes to High Serum Lipid Levels in Dutch FCHL Families and U.S. Whites With Coronary Artery Disease." *Arterioscler Thromb Vasc Biol*.

Lee, M. H., Lu, K., and Patel, S. B. (2001). "Genetic basis of sitosterolemia." *Curr Opin Lipidol* 12(2): 141-149.

Leng, G. C., Lee, A. J., Fowkes, F. G. R., Whiteman, M., Dunbar, J., Housley, E., and Ruckley, C. V. (1996). "Incidence, Natural History and Cardiovascular Events in Symptomatic and Asymptomatic Peripheral Arterial Disease in the General Population." *Int J Epidemiol* 25(6): 1172-1181.

Levy, D., Larson, M. G., Benjamin, E. J., Newton-Cheh, C., Wang, T. J., Hwang, S. J., Vasan, R. S., and Mitchell, G. F. (2007). "Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness." *BMC Med Genet* 8 Suppl 1: S3.

Lifton, R. P., Dluhy, R. G., Powers, M., Rich, G. M., Cook, S., Ulick, S., and Lalouel, J. M. (1992). "A chimaeric 11 beta-hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension." *Nature* 355(6357): 262-265.

Lin, D. Y., and Wei, L. J. (1989). "The robust inference for the Cox Proportional Hazards Model." *Journal of the American Statistical Association* 84(408): 1074-1078.

Lloyd-Jones, D. M., Nam, B. H., D'Agostino, R. B., Sr., Levy, D., Murabito, J. M., Wang, T. J., Wilson, P. W., and O'Donnell, C. J. (2004). "Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring." *Jama* 291(18): 2204-2211.

Lockhart, D. J., and Winzeler, E. A. (2000). "Genomics, gene expression and DNA arrays." *Nature* 405(6788): 827-836.

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S., and Hirschhorn, J. N. (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." *Nat Genet* 33(2): 177-182.

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., and Murray, C. J. (2006). "Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data." *Lancet* 367(9524): 1747-1757.

Lovmar, L., and Syvänen, A. C. (2006). "Multiple displacement amplification to create a long-lasting source of DNA for genetic studies." *Hum Mutat* 27(7): 603-614.

Lusis, A. J. (2000). "Atherosclerosis." *Nature* 407(6801): 233-241.

Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2007). "Entrez Gene: gene-centered information at NCBI." *Nucleic Acids Res* 35(Database issue): D26-31.

Mähonen, M., Salomaa, V., Keskimäki, I., and Moltchanov, V. (2000). "The feasibility of routine mortality and morbidity register data linkage to study the occurrence of acute coronary heart disease events in Finland. The Finnish Cardiovascular Diseases Registers (CVDR) Project." *Eur J Epidemiol* 16(8): 701-711.

Malik, S., Wong, N. D., Franklin, S. S., Kamath, T. V., L'Italien, G. J., Pio, J. R., and Williams, G. R. (2004). "Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults." *Circulation* 110(10): 1245-1250.

Manolio, T. A., Bailey-Wilson, J. E., and Collins, F. S. (2006). "Genes, environment and the value of prospective cohort studies." *Nat Rev Genet* 7(10): 812-820.

Marchini, J., Cardon, L. R., Phillips, M. S., and Donnelly, P. (2004). "The effects of human population structure on large genetic association studies." *Nat Genet* 36(5): 512-517.

Martin, C. C., Svitek, C. A., Oeser, J. K., Henderson, E., Stein, R., and O'Brien, R. M. (2003a). "Upstream stimulatory factor (USF) and neurogenic differentiation/beta-cell E box transactivator 2 (NeuroD/BETA2) contribute to islet-specific glucose-6-phosphatase catalytic-subunit-related protein (IGRP) gene expression." *Biochem J* 371(Pt 3): 675-686.

Martin, S., Nicaud, V., Humphries, S. E., and Talmud, P. J. (2003b). "Contribution of APOA5 gene variants to plasma triglyceride determination and to the response to both fat and glucose tolerance challenges." *Biochim Biophys Acta* 1637(3): 217-225.

Matarin, M., Brown, W. M., Scholz, S., Simon-Sanchez, J., Fung, H. C., Hernandez, D., Gibbs, J. R., De Vrieze, F. W., Crews, C., Britton, A., Langefeld, C. D., Brott, T. G., Brown, R. D., Jr., Worrall, B. B., Frankel, M., Silliman, S., Case, L. D., Singleton, A., Hardy, J. A., Rich, S. S., and Meschia, J. F. (2007). "A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release." *Lancet Neurol* 6(5): 414-420.

Matthaei, J. H., Jones, O. W., Martin, R. G., and Nirenberg, M. W. (1962). "Characteristics and composition of RNA coding units." *Proc Natl Acad Sci U S A* 48: 666-677.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V., Kloos, D. U., Land, S., Lewicki-Potapov, B., Michael, H., Munch, R., Reuter, I., Rotert, S., Saxel, H., Scheer, M., Thiele, S., and Wingender, E. (2003). "TRANSFAC: transcriptional regulation, from patterns to profiles." *Nucleic Acids Res* 31(1): 374-378.

Mayer, B., Erdmann, J., and Schunkert, H. (2007). "Genetics and heritability of coronary artery disease and myocardial infarction." *Clin Res Cardiol* 96(1): 1-7.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., Hinds, D. A., Pennacchio, L. A., Tybjaerg-Hansen, A., Folsom, A. R., Boerwinkle, E., Hobbs, H. H., and Cohen, J. C. (2007). "A common allele on chromosome 9 associated with coronary heart disease." *Science* 316(5830): 1488-1491.

McQueen, M. B., Bertram, L., Rimm, E. B., Blacker, D., and Santangelo, S. L. (2003). "A QTL genome scan of the metabolic syndrome and its component traits." *BMC Genet* 4 Suppl 1: S96.

Mendelsohn, M. E., and Karas, R. H. (2005). "Molecular and cellular basis of cardiovascular gender differences." *Science* 308(5728): 1583-1587.

Mills, G. W., Avery, P. J., McCarthy, M. I., Hattersley, A. T., Levy, J. C., Hitman, G. A., Sampson, M., and Walker, M. (2004). "Heritability estimates for beta cell function and features of the insulin resistance syndrome in UK families with an increased susceptibility to type 2 diabetes." *Diabetologia* 47(4): 732-738.

Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., and Devine, S. E. (2006). "An initial map of insertion and deletion (INDEL) variation in the human genome." *Genome Res* 16(9): 1182-1190.

Mishra, G. R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., Shivakumar, K., Anuradha, N., Reddy, R., Raghavan, T. M., Menon, S., Hanumanthu, G., Gupta, M., Upendran, S., Gupta, S., Mahesh, M., Jacob, B., Mathew, P., Chatterjee, P., Arun, K. S., Sharma, S., Chandrika, K. N., Deshpande, N., Palvankar, K., Raghavnath, R., Krishnakanth, R., Karathia, H., Rekha, B., Nayak, R., Vishnupriya, G., Kumar, H. G., Nagini, M., Kumar, G. S., Jose, R., Deepthi, P., Mohan, S. S., Gandhi, T. K., Harsha, H. C., Deshpande, K. S., Sarker, M., Prasad, T. S., and Pandey, A. (2006). "Human protein reference database--2006 update." *Nucleic Acids Res* 34(Database issue): D411-414.

Morgan, T. M., Coffey, C. S., and Krumholz, H. M. (2003). "Overestimation of genetic risks owing to small sample sizes in cardiovascular studies." *Clin Genet* 64(1): 7-17.

Moskau, S., Golla, A., Grothe, C., Boes, M., Pohl, C., and Klockgether, T. (2005). "Heritability of carotid artery atherosclerotic lesions: an ultrasound study in 154 families." *Stroke* 36(1): 5-8.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." *Cold Spring Harb Symp Quant Biol* 51 Pt 1: 263-273.

Mussa, F. F., Chai, H., Wang, X., Yao, Q., Lumsden, A. B., and Chen, C. (2006). "Chlamydia pneumoniae and vascular disease: an update." *J Vasc Surg* 43(6): 1301-1307.

Nabel, E. G. (2003). "Cardiovascular disease." *N Engl J Med* 349(1): 60-72.

Naughten, E. R., Yap, S., and Mayne, P. D. (1998). "Newborn screening for homocystinuria: Irish and world experience." *Eur J Pediatr* 157 Suppl 2: S84-87.

Naukkarinen, J., Gentile, M., Soro-Paavonen, A., Saarela, J., Koistinen, H. A., Pajukanta, P., Taskinen, M. R., and Peltonen, L. (2005). "USF1 and dyslipidemias: converging evidence for a functional intronic variant." *Hum Mol Genet* 14(17): 2595-2605.

Nederhand, R. J., Droog, S., Kluft, C., Simoons, M. L., and de Maat, M. P. (2003). "Logistics and quality control for DNA sampling in large multicenter studies." *J Thromb Haemost* 1(5): 987-991.

Ng, M. C., Miyake, K., So, W. Y., Poon, E. W., Lam, V. K., Li, J. K., Cox, N. J., Bell, G. I., and Chan, J. C. (2005). "The linkage and association of the gene encoding upstream stimulatory factor 1 with type 2 diabetes and metabolic syndrome in the Chinese population." *Diabetologia* 48(10): 2018-2024.

Ng, M. C., So, W. Y., Cox, N. J., Lam, V. K., Cockram, C. S., Critchley, J. A., Bell, G. I., and Chan, J. C. (2004a). "Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25." *Diabetes* 53(6): 1609-1613.

Ng, M. C., So, W. Y., Lam, V. K., Cockram, C. S., Bell, G. I., Cox, N. J., and Chan, J. C. (2004b). "Genome-wide scan for metabolic syndrome and related quantitative traits in Hong

Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25." *Diabetes* 53(10): 2676-2683.

Nilsson-Ardnor, S., Janunger, T., Wiklund, P. G., Lackovic, K., Nilsson, A. K., Lindgren, P., Escher, S. A., Stegmayr, B., Asplund, K., and Holmberg, D. (2007). "Genome-wide linkage scan of common stroke in families from northern Sweden." *Stroke* 38(1): 34-40.

Ninomiya, J. K., L'Italien, G., Criqui, M. H., Whyte, J. L., Gamst, A., and Chen, R. S. (2004). "Association of the metabolic syndrome with history of myocardial infarction and stroke in the Third National Health and Nutrition Examination Survey." *Circulation* 109(1): 42-46.

Nirenberg, M. W., and Matthaei, J. H. (1961). "The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides." *Proc Natl Acad Sci U S A* 47: 1588-1602.

Norio, R. (2003). "The Finnish Disease Heritage III: the individual diseases." *Hum Genet* 112(5-6): 470-526.

Nowak, M., Helleboid-Chapman, A., Jakel, H., Martin, G., Duran-Sandoval, D., Staels, B., Rubin, E. M., Pennacchio, L. A., Taskinen, M. R., Fruchart-Najib, J., and Fruchart, J. C. (2005). "Insulin-mediated down-regulation of apolipoprotein A5 gene expression through the phosphatidylinositol 3-kinase pathway: role of upstream stimulatory factor." *Mol Cell Biol* 25(4): 1537-1548.

O'Connell, D. L., Heller, R. F., Roberts, D. C., Allen, J. R., Knapp, J. C., Steele, P. L., and Silove, D. (1988). "Twin study of genetic and environmental effects on lipid levels." *Genet Epidemiol* 5(5): 323-341.

O'Donnell, C. J., Cupples, L. A., D'Agostino, R. B., Fox, C. S., Hoffmann, U., Hwang, S. J., Ingellson, E., Liu, C., Murabito, J. M., Polak, J. F., Wolf, P. A., and Demissie, S. (2007). "Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study." *BMC Med Genet* 8 Suppl 1: S4.

Pajukanta, P., Cargill, M., Viitanen, L., Nuotio, I., Kareinen, A., Perola, M., Terwilliger, J. D., Kempas, E., Daly, M., Lilja, H., Rioux, J. D., Brettin, T., Viikari, J. S., Rönnemaa, T., Laakso, M., Lander, E. S., and Peltonen, L. (2000). "Two loci on chromosomes 2 and X for premature coronary heart disease identified in early- and late-settlement populations of Finland." *Am J Hum Genet* 67(6): 1481-1493.

Pajukanta, P., Lilja, H. E., Sinsheimer, J. S., Cantor, R. M., Lusis, A. J., Gentile, M., Duan, X. J., Soro-Paavonen, A., Naukkarinen, J., Saarela, J., Laakso, M., Ehnholm, C., Taskinen, M. R., and Peltonen, L. (2004). "Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1)." *Nat Genet* 36(4): 371-376.

Pajukanta, P., Nuotio, I., Terwilliger, J. D., Porkka, K. V., Ylitalo, K., Pihlajamäki, J., Suomalainen, A. J., Syvänen, A. C., Lehtimäki, T., Viikari, J. S., Laakso, M., Taskinen, M. R., Ehnholm, C., and Peltonen, L. (1998). "Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23." *Nat Genet* 18(4): 369-373.

Palmer, L. J., and Cardon, L. R. (2005). "Shaking the tree: mapping complex disease genes with linkage disequilibrium." *Lancet* 366(9492): 1223-1234.

Pastier, D., Lacorte, J. M., Chambaz, J., Cardot, P., and Ribeiro, A. (2002). "Two initiator-like elements are required for the combined activation of the human apolipoprotein C-III promoter by upstream stimulatory factor and hepatic nuclear factor-4." *J Biol Chem* 277(17): 15199-15206.

Peltonen, L., Jalanko, A., and Varilo, T. (1999). "Molecular genetics of the Finnish disease heritage." *Hum Mol Genet* 8(10): 1913-1923.

Peltonen, L., Palotie, A., and Lange, K. (2000). "Use of population isolates for mapping complex traits." *Nat Rev Genet* 1(3): 182-190.

Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T. K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H. N., Rashmi, B. P., Ramya, M. A., Zhao, Z., Chandrika, K. N., Padma, N., Harsha, H. C., Yatish, A. J., Kavitha, M. P., Menezes, M., Choudhury, D. R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S. K., Madavan, V., Joseph, A., Wong, G. W., Schiemann, W. P., Constantinescu, S. N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G. C., Dang, C. V., Garcia, J. G., Pevsner, J., Jensen, O. N., Roepstorff, P., Deshpande, K. S., Chinnaiyan, A. M., Hamosh, A., Chakravarti, A., and Pandey, A. (2003). "Development of human protein reference database as an initial platform for approaching systems biology in humans." *Genome Res* 13(10): 2363-2371.

Perucho, M., Goldfarb, M., Shimizu, K., Lama, C., Fogh, J., and Wigler, M. (1981). "Human-tumor-derived cell lines contain common and different transforming genes." *Cell* 27(3 Pt 2): 467-476.

Petes, T. D., and Botstein, D. (1977). "Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast." *Proc Natl Acad Sci U S A* 74(11): 5091-5095.

Pollex, R. L., and Hegele, R. A. (2007). "Copy number variation in the human genome and its implications for cardiovascular disease." *Circulation* 115(24): 3130-3138.

Pontius, J. U., Wagner, L., and Schuler, G. D. (2003). "UniGene: a unified view of the transcriptome." *The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information.*

Poulsen, P., Kyvik, K. O., Vaag, A., and Beck-Nielsen, H. (1999). "Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study." *Diabetologia* 42(2): 139-145.

Prentice, R. L. (1986). "A case-cohort design for epidemiologic cohort studies and disease prevention trials." *Biometrika* 73(1): 1-11.

Putt, W., Palmen, J., Nicaud, V., Tregouet, D. A., Tahri-Daizadeh, N., Flavell, D. M., Humphries, S. E., and Talmud, P. J. (2004). "Variation in USF1 shows haplotype effects, gene : gene and gene : environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II." *Hum Mol Genet* 13(15): 1587-1597.

Read, M. L., Clark, A. R., and Docherty, K. (1993). "The helix-loop-helix transcription factor USF (upstream stimulating factor) binds to a regulatory sequence of the human insulin gene enhancer." *Biochem J* 295 (Pt 1): 233-237.

Reaven, G. M. (1993). "Role of insulin resistance in human disease (syndrome X): an expanded definition." *Annu Rev Med* 44: 121-131.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J., and Lancet, D. (1997). GeneCard for USF1, 29 Oct 2006, Weizmann Institute of Science, Bioinformatics Unit and Genome Center (Rehovot, Israel).

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W., and Hurles, M. E. (2006). "Global variation in copy number in the human genome." *Nature* 444(7118): 444-454.

Reiner, A. P., Carlson, C. S., Jenny, N. S., Durda, J. P., Siscovick, D. S., Nickerson, D. A., and Tracy, R. P. (2007). "USF1 Gene Variants, Cardiovascular Risk, and Mortality in European-Americans. Analysis of Two U.S. Cohort Studies." *Arterioscler Thromb Vasc Biol*.

Ribeiro, A., Pastier, D., Kardassis, D., Chambaz, J., and Cardot, P. (1999). "Cooperative binding of upstream stimulatory factor and hepatic nuclear factor 4 drives the transcription of the human apolipoprotein A-II gene." *J Biol Chem* 274(3): 1216-1225.

Rinn, J. L., and Snyder, M. (2005). "Sexual dimorphism in mammalian gene expression." *Trends Genet* 21(5): 298-305.

Risch, N., and Merikangas, K. (1996). "The future of genetic studies of complex human diseases." *Science* 273(5281): 1516-1517.

Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." *Nature* 405(6788): 847-856.

Robinson, M. D., and Speed, T. P. (2007). "A comparison of Affymetrix gene expression arrays." *BMC Bioinformatics* 8(1): 449.

Rodriguez, C. I., Girones, N., and Fresno, M. (2003). "Cha, a basic helix-loop-helix transcription factor involved in the regulation of upstream stimulatory factor activity." *J Biol Chem* 278(44): 43135-43145.

Rook, M. S., Delach, S. M., Deyneko, G., Worlock, A., and Wolfe, J. L. (2004). "Whole genome amplification of DNA from laser capture-microdissected tissue for high-throughput single nucleotide polymorphism and short tandem repeat genotyping." *Am J Pathol* 164(1): 23-33.

Rosen, N., Chalifa-Caspi, V., Shmueli, O., Adato, A., Lapidot, M., Stampnitzky, J., Safran, M., and Lancet, D. (2003). "GeneLoc: exon-based integration of human genome maps." *Bioinformatics* 19 Suppl 1: i222-224.

Roy, A. L., Meisterernst, M., Pognonec, P., and Roeder, R. G. (1991). "Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF." *Nature* 354(6350): 245-248.

Rozen, S., and Skaletsky, H. (2000). "Primer3 on the WWW for general users and for biologist programmers." *Methods Mol Biol* 132: 365-386.

Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J. C., Deleuze, J. F., Brewer, H. B., Duverger, N., Denefle, P., and Assmann, G. (1999). "Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1." *Nat Genet* 22(4): 352-355.

Saito, T., Oishi, T., Yanai, K., Shimamoto, Y., and Fukamizu, A. (2003). "Cloning and characterization of a novel splicing isoform of USF1." *Int J Mol Med* 12(2): 161-167.

Salero, E., Gimenez, C., and Zafra, F. (2003). "Identification of a non-canonical E-box motif as a regulatory element in the proximal promoter region of the apolipoprotein E gene." *Biochem J* 370(Pt 3): 979-986.

Samani, N. J., Burton, P., Mangino, M., Ball, S. G., Balmforth, A. J., Barrett, J., Bishop, T., and Hall, A. (2005). "A genomewide linkage study of 1,933 families affected by premature coronary artery disease: The British Heart Foundation (BHF) Family Heart Study." *Am J Hum Genet* 77(6): 1011-1020.

Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R. J., Meitinger, T., Braund, P., Wichmann, H. E., Barrett, J. H., Konig, I. R., Stevens, S. E., Szymczak, S., Tregouet, D. A., Iles, M. M., Pahlke, F., Pollard, H., Lieb, W., Cambien, F., Fischer, M., Ouwehand, W., Blankenberg, S., Balmforth, A. J., Baessler, A., Ball, S. G., Strom, T. M., Braenne, I., Gieger, C., Deloukas, P., Tobin, M. D., Ziegler, A., Thompson, J. R., and Schunkert, H. (2007). "Genomewide association analysis of coronary artery disease." *N Engl J Med* 357(5): 443-453.

Sawadogo, M. (1988). "Multiple forms of the human gene-specific transcription factor USF. II. DNA binding properties and transcriptional activity of the purified HeLa USF." *J Biol Chem* 263(24): 11994-12001.

Sawadogo, M., Van Dyke, M. W., Gregor, P. D., and Roeder, R. G. (1988). "Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei." *J Biol Chem* 263(24): 11985-11993.

Schulz, U. G., Flossmann, E., and Rothwell, P. M. (2004). "Heritability of ischemic stroke in relation to age, vascular risk factors, and subtypes of incident stroke in population-based studies." *Stroke* 35(4): 819-824.

SeattleSNPs, Variation Discovery Resource.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T. C., Trask, B., Patterson, N., Zetterberg, A., and Wigler, M. (2004). "Large-scale copy number polymorphism in the human genome." *Science* 305(5683): 525-528.

Sellers, T. A., and Yates, J. R. (2003). "Review of proteomics with applications to genetic epidemiology." *Genet Epidemiol* 24(2): 83-98.

Service, S., DeYoung, J., Karayiorgou, M., Roos, J. L., Pretorious, H., Bedoya, G., Ospina, J., Ruiz-Linares, A., Macedo, A., Palha, J. A., Heutink, P., Aulchenko, Y., Oostra, B., van Duijn, C., Jarvelin, M. R., Varilo, T., Peddle, L., Rahman, P., Piras, G., Monne, M., Murray, S., Galver, L., Peltonen, L., Sabatti, C., Collins, A., and Freimer, N. (2006). "Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies." *Nat Genet* 38(5): 556-560.

Sethi, A. A., Nordestgaard, B. G., and Tybjaerg-Hansen, A. (2003). "Angiotensinogen gene polymorphism, plasma angiotensinogen, and risk of hypertension and ischemic heart disease: a meta-analysis." *Arterioscler Thromb Vasc Biol* 23(7): 1269-1275.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). "dbSNP: the NCBI database of genetic variation." *Nucleic Acids Res* 29(1): 308-311.

Shieh, B. H., Sparkes, R. S., Gaynor, R. B., and Lusis, A. J. (1993). "Localization of the gene-encoding upstream stimulatory factor (USF) to human chromosome 1q22-q23." *Genomics* 16(1): 266-268.

Shifman, S., and Darvasi, A. (2001). "The value of isolated populations." *Nat Genet* 28(4): 309-310.

Sillanpää, M. J., and Auranen, K. (2004). "Replication in genetic studies of complex traits." *Ann Hum Genet* 68(Pt 6): 646-657.

Smih, F., Rouet, P., Lucas, S., Mairal, A., Sengenes, C., Lafontan, M., Vaulont, S., Casado, M., and Langin, D. (2002). "Transcriptional regulation of adipocyte hormone-sensitive lipase by glucose." *Diabetes* 51(2): 293-300.

Song, Y., Stampfer, M. J., and Liu, S. (2004). "Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease." *Ann Intern Med* 141(2): 137-147.

Soria, L. F., Ludwig, E. H., Clarke, H. R., Vega, G. L., Grundy, S. M., and McCarthy, B. J. (1989). "Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100." *Proc Natl Acad Sci U S A* 86(2): 587-591.

Souren, N. Y., Paulussen, A. D., Loos, R. J., Gielen, M., Beunen, G., Fagard, R., Derom, C., Vlietinck, R., and Zeegers, M. P. (2007). "Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities." *Diabetologia* 50(10): 2107-2116.

Stamm, S., Riethoven, J. J., Le Texier, V., Gopalakrishnan, C., Kumanduri, V., Tang, Y., Barbosa-Morais, N. L., and Thanaraj, T. A. (2006). "ASD: a bioinformatics resource on alternative splicing." *Nucleic Acids Res* 34(Database issue): D46-55.

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). "BioGRID: a general repository for interaction datasets." *Nucleic Acids Res* 34(Database issue): D535-539.

Steemers, F. J., and Gunderson, K. L. (2007). "Whole genome genotyping technologies on the BeadArray platform." *Biotechnol J* 2(1): 41-49.

Stegmayr, B., Asplund, K., Kuulasmaa, K., Rajakangas, A. M., Thorvaldsen, P., and Tuomilehto, J. (1997). "Stroke incidence and mortality correlated to stroke risk factors in the WHO MONICA Project. An ecological study of 18 populations." *Stroke* 28(7): 1367-1374.

Storey, J. D., and Tibshirani, R. (2003). "Statistical significance for genomewide studies." *Proc Natl Acad Sci U S A* 100(16): 9440-9445.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Redon, R., Bird, C. P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavare, S., Deloukas, P., Hurles, M. E., and Dermitzakis, E. T. (2007). "Relative impact of nucleotide and copy number variation on gene expression phenotypes." *Science* 315(5813): 848-853.

Strong, J. P., Malcom, G. T., McMahan, C. A., Tracy, R. E., Newman, W. P., 3rd, Herderick, E. E., and Cornhill, J. F. (1999). "Prevalence and extent of atherosclerosis in adolescents and young adults: implications for prevention from the Pathobiological Determinants of Atherosclerosis in Youth Study." *Jama* 281(8): 727-735.

Szalai, A. J., Wu, J., Lange, E. M., McCrory, M. A., Langefeld, C. D., Williams, A., Zakharkin, S. O., George, V., Allison, D. B., Cooper, G. S., Xie, F., Fan, Z., Edberg, J. C., and Kimberly, R. P. (2005). "Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level." *Journal of Molecular Medicine* 83(6): 440-447.

Tabor, H. K., Risch, N. J., and Myers, R. M. (2002). "Candidate-gene approaches for studying complex genetic traits: practical considerations." *Nat Rev Genet* 3(5): 391-397.

Tanaka, C., Mannami, T., Kamide, K., Takiuchi, S., Kokubo, Y., Katsuya, T., Kawano, Y., Miyata, T., Ogihara, T., and Tomoike, H. (2005). "Single nucleotide polymorphisms in the interleukin-6 gene associated with blood pressure and atherosclerosis in a Japanese general population." *Hypertens Res* 28(1): 35-41.

Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjold, M., Ponder, B. A., and Tunnacliffe, A. (1992). "Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer." *Genomics* 13(3): 718-725.

Tentschert, S., Greisenegger, S., Wimmer, R., Lang, W., and Lalouschek, W. (2003). "Association of parental history of stroke with clinical parameters in patients with ischemic stroke or transient ischemic attack." *Stroke* 34(9): 2114-2119.

Thanaraj, T. A., Stamm, S., Clark, F., Riethoven, J. J., Le Texier, V., and Muilu, J. (2004). "ASD: the Alternative Splicing Database." *Nucleic Acids Res* 32(Database issue): D64-69.

Thierry-Mieg, D., Thierry-Mieg, J., Potdevin, M., and Sienkiewicz, M. Identification and functional annotation of cDNA-supported genes in higher organisms., unpublished.

Tilastokeskus (2007a). Kuolleet kuolemansyyn (54-luokkainen luokitus), iän ja sukupuolen mukaan 1986-2006.

Tilastokeskus (2007b). Väestö iän (5-v.) ja sukupuolen mukaan koko maa 1865-2006.

Tishkoff, S. A., and Verrelli, B. C. (2003). "Patterns of human genetic diversity: implications for human evolutionary history and disease." *Annu Rev Genomics Hum Genet* 4: 293-340.

Travers, M. T., Vallance, A. J., Gourlay, H. T., Gill, C. A., Klein, I., Bottema, C. B., and Barber, M. C. (2001). "Promoter I of the ovine acetyl-CoA carboxylase-alpha gene: an E-box motif at -114 in the proximal promoter binds upstream stimulatory factor (USF)-1 and USF-2 and acts as an insulin-response sequence in differentiating adipocytes." *Biochem J* 359(Pt 2): 273-284.

Tregouet, D. A., Ricard, S., Nicaud, V., Arnould, I., Soubigou, S., Rosier, M., Duverger, N., Poirier, O., Mace, S., Kee, F., Morrison, C., Denefle, P., Tiret, L., Evans, A., Deleuze, J. F., and Cambien, F. (2004). "In-depth haplotype analysis of ABCA1 gene polymorphisms in relation to plasma ApoA1 levels and myocardial infarction." *Arterioscler Thromb Vasc Biol* 24(4): 775-781.

Tunstall-Pedoe, H. (1998). "Myth and paradox of coronary risk and the menopause." *Lancet* 351(9113): 1425-1427.

Tunstall-Pedoe, H., Kuulasmaa, K., Amouyel, P., Arveiler, D., Rajakangas, A. M., and Pajak, A. (1994). "Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents." *Circulation* 90(1): 583-612.

Tuomilehto, J., Rastenyte, D., Jousilahti, P., Sarti, C., and Vartiainen, E. (1996). "Diabetes mellitus as a risk factor for death from stroke. Prospective study of the middle-aged Finnish population." *Stroke* 27(2): 210-215.

Uhl, G. R., Liu, Q. R., Drgon, T., Johnson, C., Walther, D., and Rose, J. E. (2007). "Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs." *BMC Genet* 8: 10.

"The Universal Protein Resource (UniProt)." (2007). *Nucleic Acids Res* 35(Database issue): D193-197.

van der Vleuten, G. M., Isaacs, A., Hijmans, A., van Duijn, C. M., Stalenhoef, A. F., and de Graaf, J. (2007). "The involvement of upstream stimulatory factor 1 (USF1) in Dutch patients with familial combined hyperlipidemia." *J Lipid Res* 48(1): 193-200.

Vandenplas, S., Wiid, I., Grobler-Rabie, A., Brebner, K., Ricketts, M., Wallis, G., Bester, A., Boyd, C., and Mathew, C. (1984). "Blot hybridisation analysis of genomic DNA." *J Med Genet* 21(3): 164-172.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M.,

Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). "The sequence of the human genome." *Science* 291(5507): 1304-1351.

Viollet, B., Lefrancois-Martinez, A. M., Henrion, A., Kahn, A., Raymondjean, M., and Martinez, A. (1996). "Immunochemical characterization and transacting properties of upstream stimulatory factor isoforms." *J Biol Chem* 271(3): 1405-1415.

Vionnet, N., Hani, E. H., Dupont, S., Gallina, S., Francke, S., Dotte, S., De Matos, F., Durand, E., Lepretre, F., Lecoeur, C., Gallina, P., Zekiri, L., Dina, C., and Froguel, P. (2000).

"Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24." *Am J Hum Genet* 67(6): 1470-1480.

Voors-Pette, C., and de Bruin, T. W. (2001). "Excess coronary heart disease in Familial Combined Hyperlipidemia, in relation to genetic factors and central obesity." *Atherosclerosis* 157(2): 481-489.

Wain, H. M., Bruford, E. A., Lovering, R. C., Lush, M. J., Wright, M. W., and Povey, S. (2002). "Guidelines for human gene nomenclature." *Genomics* 79(4): 464-470.

Wain, H. M., Lush, M. J., Ducluzeau, F., Khodiyar, V. K., and Povey, S. (2004). "Genew: the Human Gene Nomenclature Database, 2004 updates." *Nucleic Acids Res* 32(Database issue): D255-257.

Wang, D., and Sul, H. S. (1997). "Upstream stimulatory factor binding to the E-box at -65 is required for insulin regulation of the fatty acid synthase promoter." *J Biol Chem* 272(42): 26367-26374.

Wang, D., Yang, H., Quinones, M. J., Bulnes-Enriquez, I., Jimenez, X., De La Rosa, R., Modilevsky, T., Yu, K., Li, Y., Taylor, K. D., Hsueh, W. A., Hodis, H. N., and Rotter, J. I. (2005). "A genome-wide scan for carotid artery intima-media thickness: the Mexican-American Coronary Artery Disease family study." *Stroke* 36(3): 540-545.

Wang, L., Fan, C., Topol, S. E., Topol, E. J., and Wang, Q. (2003). "Mutation of MEF2A in an inherited disorder with features of coronary artery disease." *Science* 302(5650): 1578-1581.

Wang, Q., Rao, S., Shen, G. Q., Li, L., Moliterno, D. J., Newby, L. K., Rogers, W. J., Cannata, R., Zirzow, E., Elston, R. C., and Topol, E. J. (2004). "Premature myocardial infarction novel susceptibility locus on chromosome 1P34-36 identified by genomewide linkage analysis." *Am J Hum Genet* 74(2): 262-271.

Waters, D. D., Gordon, D., Rossouw, J. E., Cannon, R. O., 3rd, Collins, P., Herrington, D. M., Hsia, J., Langer, R., Mosca, L., Ouyang, P., Sopko, G., and Stefanick, M. L. (2004). "Women's Ischemic Syndrome Evaluation: current status and future research directions: report of the National Heart, Lung and Blood Institute workshop: October 2-4, 2002 : Section 4: lessons from hormone replacement trials." *Circulation* 109(6): e53-55.

Waterston, R. H., Lander, E. S., and Sulston, J. E. (2002). "On the sequencing of the human genome." *Proc Natl Acad Sci U S A* 99(6): 3712-3716.

Watkins, H., and Farrall, M. (2006). "Genetic susceptibility to coronary artery disease: from promise to progress." *Nat Rev Genet* 7(3): 163-173.

Watson, J. D., and Crick, F. H. (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." *Nature* 171(4356): 737-738.

Weber, J. L., and May, P. E. (1989). "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction." *Am J Hum Genet* 44(3): 388-396.

Weiss, L. A., Pan, L., Abney, M., and Ober, C. (2006). "The sex-specific genetic architecture of quantitative traits in humans." *Nat Genet* 38(2): 218-222.

Weng, L., Kavaslar, N., Ustaszewska, A., Doelle, H., Schackwitz, W., Hebert, S., Cohen, J. C., McPherson, R., and Pennacchio, L. A. (2005). "Lack of MEF2A mutations in coronary artery disease." *J Clin Invest* 115(4): 1016-1020.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L. Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Miller, V., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusova, T. A., Wagner, L., and Yaschenko, E. (2007). "Database resources of the National Center for Biotechnology Information." *Nucleic Acids Res* 35(Database issue): D5-12.

White, P. C., Mune, T., and Agarwal, A. K. (1997). "11 beta-Hydroxysteroid dehydrogenase and the syndrome of apparent mineralocorticoid excess." *Endocr Rev* 18(1): 135-156.

Whitfield, J. B., Zhu, G., Madden, P. A., Neale, M. C., Heath, A. C., and Martin, N. G. (2004). "The genetics of alcohol intake and of alcohol dependence." *Alcohol Clin Exp Res* 28(8): 1153-1160.

Wienke, A., Herskind, A. M., Christensen, K., Skytthe, A., and Yashin, A. I. (2005). "The heritability of CHD mortality in danish twins after controlling for smoking and BMI." *Twin Res Hum Genet* 8(1): 53-59.

Wright, A. F., Carothers, A. D., and Pirastu, M. (1999). "Population choice in mapping genes for complex diseases." *Nat Genet* 23(4): 397-404.

Xiang, K., Wang, Y., Zheng, T., Jia, W., Li, J., Chen, L., Shen, K., Wu, S., Lin, X., Zhang, G., Wang, C., Wang, S., Lu, H., Fang, Q., Shi, Y., Zhang, R., Xu, J., and Weng, Q. (2004). "Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24." *Diabetes* 53(1): 228-234.

Yang, X. P., Freeman, L. A., Knapper, C. L., Amar, M. J., Remaley, A., Brewer, H. B., Jr., and Santamarina-Fojo, S. (2002). "The E-box motif in the proximal ABCA1 promoter mediates transcriptional repression of the ABCA1 gene." *J Lipid Res* 43(2): 297-306.

Yap, S. (2003). "Classical homocystinuria: vascular risk and its prevention." *J Inherit Metab Dis* 26(2-3): 259-265.

Yarnell, J., McCrum, E., Rumley, A., Patterson, C., Salomaa, V., Lowe, G., and Evans, A. (2005). "Association of European population levels of thrombotic and inflammatory factors with risk of coronary heart disease: the MONICA Optional Haemostasis Study." *Eur Heart J* 26(4): 332-342; discussion 317-338.

Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigos, J., and Lisheng, L. (2004). "Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study." *Lancet* 364(9438): 937-952.

Zdravkovic, S., Wienke, A., Pedersen, N. L., Marenberg, M. E., Yashin, A. I., and De Faire, U. (2002). "Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins." *J Intern Med* 252(3): 247-254.

Zeggini, E., Damcott, C. M., Hanson, R. L., Karim, M. A., Rayner, N. W., Groves, C. J., Baier, L. J., Hale, T. C., Hattersley, A. T., Hitman, G. A., Hunt, S. E., Knowler, W. C., Mitchell, B. D., Ng, M. C., O'Connell, J. R., Pollin, T. I., Vaxillaire, M., Walker, M., Wang, X., Whittaker, P., Kunsun, X., Jia, W., Chan, J. C., Froguel, P., Deloukas, P., Shuldiner, A. R., Elbein, S. C., and McCarthy, M. I. (2006). "Variation within the gene encoding the upstream stimulatory factor 1 does not influence susceptibility to type 2 diabetes in samples from populations with replicated evidence of linkage to chromosome 1q." *Diabetes* 55(9): 2541-2548.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). "Whole genome amplification from a single cell: implications for genetic analysis." *Proc Natl Acad Sci U S A* 89(13): 5847-5851.